

THE JOURNAL OF SOCIAL PSYCHOLOGY

Political, Racial,
and Differential Psychology

EDITED BY

JOHN DEWEY
CARL MURCHISON

FLOYD H. ALLPORT
Syracuse University

FRANZ BOAS
Columbia University

TRIGANT BURROW
New York City

HAVELOCK ELLIS
London

FRANKLIN H. GIDDINGS
Columbia University

G. DAWES HICKS
University of London

A. HRDLÍČKA
Smithsonian Institution

CHARLES H. JUDD
University of Chicago

DAVID KATZ
Universität Rostock

K. KORNILOV
Gosudarstvennii Institut
Experimentalnoi
Psikhologii, Moskva

L. LÉVY BRUHL
Université de Paris

MARK A. MAY
Yale University

WILLIAM McDUGALL
Duke University

RICHARD MÜLLER-FREIENFELS
Pädagog. Akademie, Stettin

RAYMOND PEARL
The Johns Hopkins University

WALTER B. PILLSBURY
University of Michigan

A. T. POFFENBERGER
Columbia University

BERTRAND RUSSELL
Petersfield, England

EDWARD SAPIR
University of Chicago

CARL E. SEASHORE
University of Iowa

GIUSEPPE SERGI
R. Università di Roma

L. L. THURSTONE
University of Chicago

CLARK WISSLER
Yale University

KIMBALL YOUNG
University of Wisconsin

LUBERTA M. HARDEN, PH.D.
Assistant Editor

VOLUME III

1932

Published quarterly by the
CLARK UNIVERSITY PRESS
Worcester, Massachusetts
U. S. A.

THE JOURNAL OF SOCIAL PSYCHOLOGY

Editorial Board
and Editorial Psychology

EDITED BY
JOHN DEWEY
CARL MURCHISON

Copyright, 1932,

by

CLARK UNIVERSITY

VOLUME III
1932

Published monthly by the
CLARK UNIVERSITY PRESS
Weston, Massachusetts
U. S. A.

TABLE OF CONTENTS

HOWARD BECKER AND DAVID K. BRUNER	
Some Aspects of Taboo and Totemism	337-353
HANNAH M. BOOK	
A Psychophysiological Analysis of Sex Differences . . .	434-462
DAVID K. BRUNER AND HOWARD BECKER	
Some Aspects of Taboo and Totemism	337-353
W. D. COMMINS	
The Marriage-Age of Oldest Sons	487-490
C. E. ELLISON, H. H. REMMERS, AND A. J. HARRIS	
The Relation between Liberal and Conservative Attitudes in College Students, and Other Factors	320-336
G. C. FERRARI	
The Problem of Criminality in Children	66-78
EDWIN G. FLEMMING	
Pleasing Personality	100-107
EDWIN G. FLEMMING	
Testing Some Aspects of Personality	376-384
EDWIN G. FLEMMING	
Best Friends	385-390
PAUL HANLY FURFEY AND J. EDWARD RAUTH	
Developmental Age and Adolescence	469-472
K. C. GARRISON	
The Relative Influence of Intelligence and Socio-cultural Status upon the Information Possessed by First-Grade Chil- dren	362-367
K. C. GARRISON	
A Study of the Emotional Characteristics of Women Prison- ers	472-477
EDWARD B. GREENE	
The Michigan Non-verbal Series	354-359
A. J. HARRIS, H. H. REMMERS, AND C. E. ELLISON	
The Relation between Liberal and Conservative Attitudes in College Students, and Other Factors	320-336
O. L. HARVEY	
Some Statistics Derived from Recent Questionnaire Studies Relative to Human Sexual Behavior	97-100
O. L. HARVEY	
The Scientific Study of Human Sexual Behavior	161-188
O. L. HARVEY	
Concerning the Thurstone "Personality Schedule" . . .	240-251
O. L. HARVEY	
Concerning the Thurstone "Personality Schedule"—Erratum	354
KATHARINE WOOD HATTENDORF	
A Study of the Questions of Young Children Concerning Sex: A Phase of an Experimental Approach to Parent Education	37-65

JOSEPH WYMAN HAWTHORNE	
A Group Test for the Measurement of Cruelty-Compassion: A Proposed Means of Recognizing Potential Criminality	189-211
KATE HEVNER	
A Method of Correcting for Guessing in True-false Tests and Empirical Evidence in Support of It	359-362
E. D. HINCKLEY	
The Influence of Individual Opinion on Construction of an Attitude Scale	283-296
LETA S. HOLLINGWORTH AND RICHARD MADDEN	
How One Race Judges Another for Physical Attractiveness	463-469
JAMES QUINTER HOLSOPE	
The Social Adjustment of Delinquents Who Are Unable to Inhibit Old Automatic Perceptual Responses	91-96
GRANVILLE B. JOHNSON	
Sex Differences in Learning to Walk a Tight Wire	114-118
VERNON JONES	
Relation of Economic Depression to Delinquency, Crime, and Drunkenness in Massachusetts	259-282
CHARLES C. JOSEY AND CARROLL H. MILLER	
Race, Sex, and Class Differences in Ability to Endure Pain	374-376
HARVEY C. LEHMAN AND PAUL A. WITTY	
Nervous Instability and Genius: Military and Political Leaders	212-234
CLARENCE LEUBA	
The Measurement of Incentives and Their Effect: A Contri- bution to Methodology and Orientation Resulting from the Experimental Use of Incentives	107-114
RICHARD MADDEN AND LETA S. HOLLINGWORTH	
How One Race Judges Another for Physical Attractiveness	463-469
MARK A. MAY	
Problems of Measuring Character and Personality	131-145
CARROLL H. MILLER AND CHARLES C. JOSEY	
Race, Sex, and Class Differences in Ability to Endure Pain	374-376
RUDOLF PINTNER	
The Influence of Language Background on Intelligence Tests	235-240
KARL C. PRATT	
A Note upon the Relation of Activity to Sex and Race in Young Infants	118-120
J. EDWARD RAUTH AND PAUL HANLY FURFEY	
Developmental Age and Adolescence	469-472
H. H. REMMERS, A. J. HARRIS, AND C. E. ELLISON	
The Relation between Liberal and Conservative Attitudes in College Students, and Other Factors	320-336
BELLE SCHILLER	
A Quantitative Analysis of Marriage Selection in a Small Group	297-319
ROSS STAGNER	
Differential Factors in the Testing of Personality: I. Sex Differences	477-482
ROSS STAGNER	
Differential Factors in the Testing of Personality: II. Differ- ences in Maturity	482-487

L. L. THURSTONE	
The Isolation of Blocs in a Legislative Body by the Voting Records of Its Members	425-433
CHARLES K. A. WANG	
A Scale for Measuring Persistence	79-90
CHARLES K. A. WANG	
Suggested Criteria for Writing Attitude Statements . . .	367-373
RAYMOND R. WILLOUGHBY	
A Scale of Emotional Maturity	3-36
RAYMOND R. WILLOUGHBY	
The Functions of Conversation	146-160
RAYMOND ROYCE WILLOUGHBY	
Some Properties of the Thurstone Personality Schedule and a Suggested Revision	401-424
PAUL A. WITTY AND HARVEY C. LEHMAN	
Nervous Instability and Genius: Military and Political Leaders	212-234
Announcement of Trips to German and Austrian Psychological Institutes at the Conclusion of the International Congress of Psychology in Copenhagen, August 22-27, 1932	252

Books

LURA BEAM AND ROBERT LATOU DICKINSON	
A Thousand Marriages: A Medical Study of Sex Adjust- ment (Raymond R. Willoughby)	491-493
ROBERT LATOU DICKINSON AND LURA BEAM	
A Thousand Marriages: A Medical Study of Sex Adjust- ment (Raymond R. Willoughby)	491-493
CARL MURCHISON [ED.]	
A History of Psychology in Autobiography, Volume I, and Psychologies of 1930 (R. M. Ogden)	121-128
P. M. SYMONDS	
Diagnosing Personality and Conduct (Gordon W. Allport)	391-396
LEE EDWARD TRAVIS	
Speech Pathology: A Dynamic Neurological Treatment of Normal Speech and Speech Deviations (R. H. Stetson) . .	253-255
DOROTHY HAZELTINE YATES	
Psychological Racketeers (Lewis M. Terman)	493

A SCALE OF EMOTIONAL MATURITY*

RAYMOND R. WILLOUGHBY

ORIENTATION AND CONCEPTS

There is much that is unsatisfactory about the concept of emotional maturity. It is clinical, ultimately psychoanalytic, and therefore controversial; it is difficult to define clearly; it is not certain that the noun is not a misnomer; and it baffles analysis of the satisfactory kind that permits the construction of tests for each component or condition. Meanwhile, the accumulating weight of recorded observation forces a belief that there is an entity or group of entities in personality, closely correlated with capacity for happy, full, and effective living, which consists essentially in a loosening and slipping away of attitudes and interests which are tolerable in children, but fatal in adults; these attitudes appear to consist in an over-preoccupation with the self and its satisfactions, a too great absorption of the field of attention with the ego. It is entirely possible that the implication of longitudinal development is a mistake; it may be that the syndrome in question has more affinities with the course of an illness than with normal growth—that it may be a sort of metabolic tension and release. At any rate, the facts that a high degree of emotional “maturity” is occasionally seen in children who have been psychologically handled, and that the overwhelming majority of adults are plainly “immature” in this respect (and the older probably more so than the younger) render simplistic the suggestion twice offered the present investigator to the effect that the correct methodology is to observe the emotional traits which alter or develop with age. The term may have to be abandoned altogether, but in any case it cannot be understood so literally. The study cannot be grounded upon the statistical concept of normality—cannot, to be useful, be a mere description of the *status quo*—because there is beginning to be evidence that the statistical normal represents a kind of widespread arrest of development.

*Accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office, October 31, 1930.

Neither, *a priori*, can a test (in which the subject himself reacts to the experimental situation) be regarded as satisfactory for the estimation of this elusive quality (the empirical findings on the use of the present scale as a test are presented below): the need to see oneself as the sort of individual approvable by the group is quite capable of distorting the estimate beyond the bounds of usefulness. It is necessary that the estimate be made by others without the subject's knowledge—that is, by means of a scale. And since the raters using such a scale in a non-experimental situation will not ordinarily be skilled in the evaluation of the clinical signs of emotional maturity, it is necessary that the expert judgments concerned be embodied in the scoring key, while the work of the rater be confined to determining whether the subject has or has not the sign.

Such a scale, if technically satisfactory, should be of practical as well as theoretical interest. If no person were selected as policeman, foreman, salesman, minister, army officer, unless he had attained a high degree of emotional maturity, a great reduction would probably ensue in the sum of the frictions of daily life; while if the same criterion were demanded of teachers and stressed for parents, the possibilities for improvement are incalculable.

The definition used in the present study is due to Abraham; other suggestive writers on the concept include Freud, Holt, Krutch, and Lippman. The treatment of data is patterned upon that of Thurstone.

The dialectic of the present study is simple: 150 items were formulated (see "Scale") which were believed to be diagnostic of emotional maturity; they were submitted to competent judges (see "Judges") who estimated the degree of emotional maturity of which they were diagnostic; these judgments were pooled to derive score values. Blanks containing the same items were then submitted to raters, two of whom judged the same subject (the subject also submitted a self-rating); these results were analyzed item by item (chiefly for reliability of rating, *i.e.*, agreement between raters, and discriminatory value of the item), and on the basis of this analysis the scale was reduced to 60 items and a scoring system derived.

The following is the definition submitted to the judges:

Emotional maturity, in the understanding of the investigator, is freedom from narcissism and from ambivalence; in other terminology, it is release from egocentrism, the achievement of socialized impulses, of insight; emotional acceptance of the

reality principle and an "analyzed" condition are also approximate synonyms. If you find yourself in serious disagreement with this concept, please check here (.....) and write your own definition on the reverse side of this card, *but rate according to the definition given above.*

There were two disagreements with the definition (2%). One judge found it "intolerably psychoanalytic," the other objected to the terms narcissism and ambivalence on the ground that they were derived from mythology and chemistry respectively, and recommended that the investigator stick to observed facts. Neither offered a substitute definition.

CRITICISMS

The principal criticisms fall into a few main classes: (a) The whole project is too psychoanalytic in flavor. There seems to be no way of meeting this objection satisfactorily; insofar as it is not due to the investigator's impolitic handling of the terminology, it appears to rest on the objectors' affective organization, and, therefore, to be unalterable. (b) The items reflect too greatly the investigator's own biases. Several of the offending items have been eliminated on the ground of unsatisfactory statistical behavior; the terminology of others has been reexamined and altered (to the extent that such procedure is believed not to interfere with the sense as judged and rated) in the administrative form of the scale. (c) The language is too involved. Reexamination and alteration have partly met this objection; but nothing fundamental can be done about it, since the maintenance of a satisfactory incidence demands a considerable degree of abstractness in the terminology. While the scale is not intended to be restricted to use by psychiatrists, it cannot be used by naïve or uneducated persons. (d) A high degree of emotional maturity as judged by the scale appears to connote absence rather than maturity of emotions. To some extent this appearance may be ascribed to the judges, who were free to place items diagnostic of absence of emotion in a lower group if they wished; but they were limited, of course, by the structure of the items given them to judge. A long discussion on fundamental concepts could no doubt be indulged on this topic, but it would probably be unfruitful for lack of supporting data; it is sufficient to say that it is in fact an open question whether the presence of the cataclysmic states usually thought of as emotions (fear, love, anger, etc.) is not indeed to be thought of as indicating incomplete

psychical development. Much might be said for the proposition that *Sturm und Drang* are essentially adolescent characteristics, and that the fully developed personality is quiet, smoothly creative, and relatively free from exaltations and despairs alike. 250

METHOD OF JUDGING

The process of estimating the amount of emotional maturity of which each item may be regarded as diagnostic (the score or value of the item) was to submit the 150 items, each printed upon a 3 x 5 card, in random order (the same for each judge), with the following instructions:

On each of the following cards a situation and a response thereto are stated in terms of *S*, a subject; the judgments desired are of the degree of emotional maturity in *S* indicated by this response to this situation. These judgments are to be effected by (1) dividing the cards into three groups, high, medium and low, and (2) dividing each of these groups in turn into high, medium and low, making nine groups in all. Group 1 will represent the greatest degree of emotional maturity practically attainable in our civilization; Group 9 will represent the lowest degree of emotional maturity compatible with the continued functioning of the individual in the social group. The intervals between groups are to be as nearly equal as possible; inequalities in size of groups are to be disregarded. Head cards are furnished to separate the groups.

This process is herein referred to as judging, and the persons who did it are called judges, to distinguish them from the raters, who rated actual subjects for the presence or absence of the traits indicated by the items.

Ogive curves were constructed for the 101 judgments received, and the median (score or value) and semi-interquartile range (spread) estimated graphically (Thurstone's method of equal-appearing intervals) to the nearest tenth of an inter-group interval. Extrapolation was resorted to where necessary, the resulting values giving every indication of being reasonably well determined within the probable error of the points (about 6%).

JUDGES

The judges (i.e., those who estimated the amounts of emotional maturity of which the items were diagnostic) were 101 in number, selected primarily from workers in child guidance clinics and supple-

mented by teachers of psychology and interested and competent laymen. Descriptive statistics are available for about 97 (exact figures different for the different topics).

Eleven (9 men, 2 women) have been psychoanalyzed, and 3 more (men) answer "?," "By self" and "Sort of." The numbers of hours spent in analysis were "Post-graduate study and year" (presumably about 225), 120, 120, 100, "3 mos." (presumably about 75), 40, 30, 30, 30, 20, 20, 15, 10, and 7.

The highest academic degrees received were: Ph.D., 28; M.D., 24; A.M., M.S., M.S.S., or other master's, 17; B.A., B.S., 19 (includes several who have done graduate work); normal school graduate, 1; unknown, 8.

The age distribution is:

	20	25	30	35	40	45	50	55	60	Total
Men	1	10	13	15	9	5	3	1		57
Women		7	4	8	3	1		4	1	28
Both	1	17	17	23	12	6	3	5	1	85

For evaluating the answers to requests for a brief self-estimate of adequacy of social, sexual, and religious adjustment, a grouping of *superior*, *average*, and *inferior* has been adopted. The first comprises such responses as "adequate," "good" (unless accompanied by "excellent"), 1, 2, or 3 where the groups of the judgment were used to characterize adjustment, "gets along well with people," "95%," "Satisfactory," "Normal," "No conflict," "Active," "High average"; as *average* are classified such responses as "fair" (unless accompanied by "good"), 4, 5, and 6 in terms of groups, "resigned," "average," "unmarried—no repulsions or craving" (sex); under *inferior* are included responses like "poor," "fair only," 7, 8, or 9 on the basis of groups, "average or slightly below," "some conflict"; the religious adjustment is somewhat difficult of evaluation, but the criterion has been apparent psychic comfort on any basis, rather than conformity or non-conformity. The distribution is as follows:

	Superior	Average	Inferior	Total
Social	41	27	14	82
Sexual	42	33	7	82
Religious	48	26	8	82

The quantitative aspects of the experience with personality problems may be expressed in terms of the total number of years of experience where this figure was mentioned:

0	1	5	10	15	20	25	Total	Median
3	13	17	7	6	1	2	49	7.5

A sampling of every fourth case at random will indicate the qualitative aspects:

8 years clinical work in psychiatry, 3 years executive work.

Psychiatric social worker 7 years.

Very little.

Psychopathic hospital, jail surveys, problem clinic for children.

17 years in clinical psychiatry, examination of mental defectives and criminals.

Parent of 4 children, member of legal profession.

Has worked with a few mental patients.

8 years psychiatric social work, including 2 years child guidance clinic.

18 years teaching, work with student adjustment problems.

High school principal, 2 years clinical assistant in university.

Interest in work with problematic personality.

Vocational guidance, clinical work, college teaching.

Neurology and psychiatry for 20 years.

17 years in state hospital service.

Personnel work in college.

14 years psychiatric work.

None.

Psychopathology and psychotherapy 3 years.

4 years clinical experience.

3 years.

Worked with Prof. G. W. Allport; experiments under Prof. F. H. Allport.

Yes.

Is principal interest; research in field.

To these persons, many of whom showed unusual interest in the project and took time out of busy programs to do extra work, to secure other collaborators, and to write helpful letters of suggestion, the deep gratitude of the investigator is sincerely expressed; it is regretted that various considerations make a more personal acknowledgment impossible.

VALUES

The scores, or emotional maturity values (medians for 101 judges), for the original and final (selected) scales, in terms of tenths of inter-group intervals, are as follows (index is lower class limit, and negative values are due to extrapolation):

	-30	-20	-10	0	10	20	30	40	50	60	70	80	Total
Original	1	1	4	19	6	3	9	21	24	29	22	11	150
Final	1	1	2	12	1	3	7	10	12	9	2		60

The derived scores (reversed in order to assign a high value to a high degree of emotional maturity) distribute as follows:

9	8	7	6	5	4	3	2	1	Total
2	6	8	3	3	14	11	9	4	60

SPREAD

The semi-interquartile range of judgments as to the degree of emotional maturity indicated by the items ("spread") distributes as follows for the original and final scales (index is lower class limit; tenths of inter-group intervals):

	2	4	6	8	10	12	14	16	18	...	24	Total
Original	1	13	20	28	47	22	11	5	2	...	1	150
Final		5	6	13	19	9	8					60

CONSISTENCY OF JUDGMENT

Due to the kindness and interest of Mrs. Theresa Kennedy of Long Beach, California, it is possible to present some data on the consistency of individual judgment. Mrs. Kennedy judged the cards independently on October 6, 1929, December 8, 1929, and March 7, 1930. Her results may be outlined as follows: defining "up" as a movement toward increased emotional maturity (as Group 9, Group 8, Group 6 for the three judgments in order, or 996, or 877) "up-down" as a movement first up, then down (as 867), the distribution is as follows:

Up	Up-Down	No Change	Down-Up	Down	Total
74 (36 996's)	12	29	14	21	150

The distribution of mean deviations from the mean (in groups) is:

0	1	2	3	4	5	6	7	8	9	Total
29	15	7	18	50	22	5	1	1	2	150

Dividing the groups into thirds (A, Groups 1-3; B, Groups 4-6; C, Groups 7-9), the movement of items can be traced through the three judgments:

Judgment 1	Judgment 2	Judgment 3
A - 43	{ A - 39 B - 2 C - 2	{ A - 35 B - 4 C - 1 A - 1 B - 1 C - 1 A - 2 B - 2 C - 2
B - 26	{ A - 3 B - 14 C - 9	{ A - 3 B - 2 C - 10 A - 1 B - 3 C - 5
C - 81	{ A - 1 B - 9 C - 71	{ A - 1 B - 1 C - 1 A - 5 B - 4 C - 62 A - 9 B - 9 C - 9
Total A	43	42
B	26	72
C	81	36
	150	150

There is a systematic displacement, for this judge, amounting, on the average, to approximately two-thirds of a group interval "upward" for each item from first to last judgment; this is largely due

to the frequency of the pattern CCB, and particularly to that of the pattern 996. In connection with this individual result it may be mentioned that the probable error of a point on the ogive curves from which the e.m. values were derived is about 6%; these facts have a bearing upon the apparent liberties taken in converting the obtained values into derived scores for greater ease in scoring.

ADDITIONAL ITEMS

In response to a request to the judges to submit additional items, particularly items indicative of a high average emotional maturity, 67 items were received. It was not feasible to have these tested statistically as the preceding 150 had been, but they were judged by four of the "best" judges found in the preliminary investigations ("best" in the sense that their judgments conformed closely to the average judgments and had a narrow spread) and in 53 cases by the authors also. The means of these four or five judgments distribute as follows:

10-	15-	20	25	30	35	40	45	50	55	60	65	70	75	80-	85-	Total
12	11	10	1	4	3	2	3	1	1	3	4	4	1	5	2	67

The spread is not great; it may be indicated by giving the judgments for every fourth item at random:

5454	77765	6966	86784
21321	87688	47648	42321
83651	22321	31221	53532
21131	98897	43521	21243

Although several of the items are probably too specific and personal (i.e., have too low an incidence, as defined below) to be usable, and others are nearly duplicates, all are given herewith verbatim, with the respective mean judgments in tenths of an inter-group interval:

Ability to be sincerely, not sardonically, amused at one's own behavior patterns. (10)

S has a sense of humor which permits others to laugh at the things he loves and still to love them. (10)

S is faced by the possibility of a prolonged and painful illness, which can only be fatal in its outcome, of one affectively near to him. He accepts the situation and readjusts his life accordingly. (10)

The wife of *S*, *whom he loves*, confesses love for another man; *S* talks it over with her and is agreeable to a legal divorce. (12)

S spontaneously laughs with others at any joke on himself. (12)

Upon the death of one affectively near to *S*, he accepts it as requiring the rearrangement of his interests and hopes, but with a minimum of grief. (12)

S seems to enjoy a joke on himself as well as anyone else. He laughs about it without being loud or self-conscious about it. (12)

S is in a situation the outcome of which is greatly affect-laden; it is similar to a previous experience in which his handling of the issues resulted disastrously. He now selects the elements of the old situation which can be of advantage at present, but avoids making the same errors. (12)

S's ambitions are in harmony with his capacities. His limitations are accepted by himself without emotion. He enjoys making the most of his capacities, which he also accepts realistically. (In a word, he has good insight regarding himself.) (12)

S is conscious of attentions to his spouse on the part of a man of apparently superior parts. *S* takes the long view of things, intending that the best man shall win. (13)

S is actively interested in divergent opinions and tries to investigate their genesis and history. (14)

S accepts responsibility. Sexually, he marries, has children and provides adequately for the family. Socially, he considers the group as well as himself and does not seek to escape the burdens of his work. (14)

S enjoys games, amusements, exercise, etc., but is not dominated by any one or by all. They serve to recreate him (rather than divert). They do not interfere with his work, but rather add zest to his main activity. (16)

S discovers he has been betrayed by a friend, changes his behavior toward other friends that they may not similarly impose on him; accepts his betrayal as in part due to his own reactions, and continues to have confidence in remaining friends. (16)

S is able to remark another's shortcomings without excitement or taking sides (or condemning the person in toto). (16)

S reacts to disparagement of his personal qualities by taking a philosophically detached attitude toward the disparager. (16)

S arrives at the retiring age calmly, accepting his waning powers, and endeavors to enjoy the rest of life. (17)

S recognizes the objective defects of his love object in comparison with objects of his friends, without subsequent decrease in or ambivalence of his affection for her. (18)

S finds the work he has chosen in life agreeable and derives a good share of his satisfaction in doing his work well. (18)

S and wife discuss up-to-date topics at the dinner table; contributions by their two children are solicited and treated with respect. (18)

S welcomes opportunities for enjoying good health, good companionship and good sport, with no carry-over of sorrows or painful experiences. (18)

S, accepting modern dynamic psychology, discusses religion calmly and dispassionately. (18)

S does not permit attachments to either parent to come between his immediate family and himself. (18)

S has a depression following a severe operation, but refuses to pass judgment on himself until the depression lifts. (20)

S enjoys competition in his work or in games, and does not gloat over his successes or play the "poor sport" when he loses. (20)

S takes pride in the success of his immediate associates and encourages and stimulates those younger than himself. (20)

S possesses a high degree of idealism not only in connection with personal religious attitudes but also in social situations. This does not interfere with an ability to face reality. Both acceptance of real situations and idealistic attitudes are included in final judgment. (21)

S feels that the purpose of art is to elucidate life; that a just and clear representation of the less beautiful sides of life is admissible as art. (22)

S is in a group in which the conversation is general. He takes part in this, refraining from conversing with a friend, also present, on a topic of interest only to them both, until a lull or other opportunity occurs. (24)

S recognizes the objective defects of his love object in comparison with objects of his friends; and then compares self with friends. Concludes it is a two-sided situation and accepts it in good part. (24)

S marries or has child. *S* takes out insurance which he carries without pride, self-pity, or feeling of loss (of money spent); and without reference to it in presence of family in terms either of joking, evidence of affection, or burden it imposes. (24)

In bereavement or disaster *S* shifts his attention to the assistance of others in distress. (24)

S expressed willingness to quit college until her father recovered from a business failure. (24)

S has decisive views upon political situations but does not express these views except where they will be received constructively. (26)

S derives esthetic enjoyment from various arts, music, literature, drama, painting, etc. (30)

S has hobbies from which he derives satisfaction supplementary to that from his main activity. (30)

S does not object to publicity for his attainments but dislikes and avoids it for himself. (33)

S eats at the college commons instead of at his fraternity, in order that he may better appreciate the general student point of view. (34)

S spends his leisure time in the standard amusements for his age, sex and social group. (35)

S has a well worked out philosophy of life and sticks to it excepting for unavoidable moments and necessary distractions. (36)

S feels he has a legitimate paternal antagonism and retains his attitude without self-reproach even when his father dies. (37)

S is always serious-minded, and inclines to be self-analytical but not needlessly worried. (42)

S has a marked physical defect, but endeavors to make up for this by dressing very well. (42)

S is frankly complimented (i.e., praised in proportion to his merits); he is pleased and emotionally touched. (45)

S orders all his acts under a comprehensive religious plan. Sin to him is deviation from his life purpose, which he conceives to be a line of conduct formed by him in relation to the plans of the Deity as he understands them. (46)

S is rather neglectful of his health, giving special attention to physical matters only when discomfort arises and persists. (47)

S is obviously flattered (i.e., praised to a greater degree than his performance merits); he manifests more or less simultaneous pleasure at the praise and skepticism as to its genuineness. (53)

S has no interest in government and prefers to leave its direction to others. (58)

S cannot discuss another without making comparisons with himself. (60)

S almost invariably expresses a like or dislike for a person who is being discussed. (62)

Tends to judge individuals or evaluate situations according to personal comfort or discomfort felt with regard to them. (64)

S's family make much of every birthday and anniversary. Feelings are

hurt if one is unavoidably away or if a special day is forgotten. (66)

S possesses a high degree of idealism not only in connection with religious attitudes but also in social situations. Is afraid to compare this with reality. (68)

S has decisive views upon political situations and feels that for the good of his country he should make others agree with him. (68)

S ordinarily reacts to situations comprehensively, but where affect is involved has a tendency to react to the affective element in isolation. (68)

S is exposed to unsocial behavior on the part of an individual with whom he is closely associated and consistently withdraws from conversation in the presence of this person, not because of a desire to do so but because he feels something akin to fear which prevents a normal flow of thought and speech. (70)

S's sense of humor is dominated by *Schadenfreude*. (72)

S thinks of marriage as sacred and divorce as wrong. Even if love ceases the marriage tie is considered indissoluble. (72)

S cannot keep his mind off his business, so that at home he is morose and distant. (74)

S has never married, as his middle-aged widowed mother has always persuaded him that it was his duty to live with her as long as she lived. (76)

S has violent likes and dislikes. Certain topics are taboo in his presence. Has to be calmed, pacified or dragged off! (82)

S, although married and living in the East, insists that his family spend all of their yearly vacation in making a trip to the Middle West to visit his parents. (83)

S is in a situation the outcome of which is greatly affect-laden; it is similar to a previous experience in which his handling of the issues resulted disastrously. He now avoids the situation entirely, being emotionally upset by any reminder of the earlier disaster. (83)

S had few educational advantages when young, but in middle life managed to complete a business course. She is constantly being thrown in contact with cultured people whom she never misses an opportunity to annoy even tho she tries to imitate them. (84)

S is agreeable and pleasant at the office so long as he is consulted on all questions and can dictate in departments other than his own. When not deferred to he begins a personal persecution of the individual or group not consulting him. (84)

S is in a situation the outcome of which is greatly affect-laden; it is similar to a previous experience in which his handling of the issues resulted disastrously. He now repeats his behavior of the earlier experience. (85)

S is a physician whose position in an institution makes a woman medical assistant necessary. He asserts that women do not make capable doctors and makes the position a very trying one for each of his many assistants, who follow each other in rapid succession. (88)

The following items were submitted in questionnaire form, but were not judged:

Are you bothered when you feel the effect of your appearance on others is not what it might be? (For instance, poorly dressed.)

If a person wrongs you, can you easily forget it, or do you always remember it when you are with that person?

Do bothersome things quickly excite you, and do you quickly get over it?

Are you often angered? At people or things?

Are you considered easily led?

Are you terribly annoyed when you are compelled to wait for a slow person? (For instance, a clerk wrapping a package, or a partner hesitating at bridge.)

Are you quick to sympathize with one who brings a hard-luck story?

Do you feel a hatred for anyone which you feel unable to control?

RATING METHOD

The items were printed, in the same (chance) order in which they were submitted to the judges, on a four-page folder and submitted to about 250 college students of both sexes, with the request to rate themselves and to secure independent ratings from two persons well acquainted with them.¹ Each item was followed by the digits 0 1 2 3; the instructions were:

"S" in each of the following situations is, for the moment, your subject. If you have observed him in a situation much like that described, and noted that his reaction was (a) similar to the one described, circle the 3; (b) different from the one described, circle the 0. If you have never observed him in a situation like that described, but believe that his reaction to it would be (a) similar to the one described, circle the 2; (b) different from the one described, circle the 1.

There is no evidence that any difficulty was encountered by any of the subjects in comprehending these directions; but there is considerable evidence that perhaps a quarter of the subjects found the phraseology of the items unfamiliar and difficult. The difficulty, however, appeared to be of a local nature, as in some groups large numbers of forms were correctly executed without complaint, while in others even extensive explanation on the part of instructors seemed of doubtful clarifying power.

DISAGREEMENT

Items which yield a high percentage of cases in which the two raters disagree (0 or 1 vs. 2 or 3) as to the presence of the trait in their subject are evidently undesirable, since they tend to cut down

¹For indispensable assistance in securing the cooperation of these subjects (to whom also the investigator acknowledges his great indebtedness), thanks are due to Drs. V. Jones of Clark University, D. E. Johannsen of Wellesley College, P. H. Ewert of the University of Vermont, A. H. Estabrook of Colgate University, H. Faterson of Smith College, M. B. Jensen of Michigan Central State Teachers' College, F. A. Moss of George Washington University, G. W. Allport of Syracuse University, F. L. Goodenough of the University of Minnesota, F. H. Allport of Dartmouth College, N. L. Munn of the University of Pittsburgh, and P. E. Vernon of Yale University.

reliability of rating. The following distributions show the percentages of disagreement for the original and final forms of the scale (index is lower class limit):

	5	10	15	20	25	30	35	40	45	Total
Original	5	11	14	23	20	41	20	14	2	150
Final				1	4	29	17	9		60

The reason for the more drastic elimination from the more favorable end of the distribution appears in the section on "Incidence."

INCIDENCE

A low percentage of disagreement, at first glance apparently favorable, may arise from an unusually high or unusually low incidence of the trait in question; that is, there may be high agreement based upon the almost universal presence or absence of the trait. Taking into account this consideration, an item will be acceptable in proportion as its incidence approaches 50%—a criterion which will be seen to operate in the reverse direction from that of the percentage of disagreement. In these computations a trait was regarded as present when both raters marked it positively, the self-rating being disregarded. A direct measure of incidence cannot, however, be derived from the crude percentage of positive agreement, since the effect of the disagreements would not be considered thereby; accordingly, the measure taken, which may be called the incidence, was the percentage of the balance left by the removal of the disagreements constituted

by the positive agreements $\left(\frac{P}{100-D} \right)$. Thus, an item with 30%

disagreements and 35% positive agreements was credited with an incidence of 50. The most extensive eliminations were due to the application of this criterion; the following distribution gives the incidences in the original and final scales (index is lower class limit):

	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Total
Original	16	31	10	12	16	4	4	8	2	4	4	5	1	4	7	3	5	4	6	4	150
Final				8	9	4	4	6	2	3	1	4	1	3	7	3	5				60

Plotting incidence against disagreement for the original scale gives a curious crescentic distribution, very well marked, convex toward high disagreement and the horns (in low disagreement) at high and

low incidence respectively; a phenomenon difficult to explain until it is recalled that items inherently difficult to judge will yield both a high percentage of disagreements between raters on the same subject, which is high disagreement as here defined, and a high (approaching chance) percentage of opposite judgments as between pairs of judges on different subjects, which is medium incidence as here defined.

SCORING

The problem of scoring may be resolved into three sub-problems.

1. Whether the scores in terms of group intervals of judgment, just as derived from the pooling of the judgments, are sufficiently superior to some less cumbersome derivative to warrant the extra labor involved in their use. On the basis of a trial population of 70, the reliability of rating for the median scores (see point 3 below) using the group-interval scores is $.53 \pm .06$; using a derivative (see below) it is $.565 \pm .055$. In terms of this criterion, therefore, there is no practical advantage in preserving the original scores.

2. Whether a distinction should be made between the ratings of 2 and of 3, and, if so, what combination should be used. Under the heading "Degree Ratings" it will be shown that both 2's and 3's should be used without distinction.

3. Whether the mean or the median of the item values for a given subject should be used as his score. There are advantages of convenience in favor of the mean (since it is more readily understood and does not require the construction of a frequency distribution) which should be utilized if they are not outweighed. On the basis of the trial population of 70, the reliability of rating of the derivative scores, using the means, is $.544 \pm .06$; using the medians, it is $.565 \pm .055$. The standard error of this difference of .02 may be estimated at .10; calculated without the third term under the radical in the formula for the standard error of a difference, it is .12. There are no significant differences between the central tendencies or variabilities of the two kinds of scores. The mean may accordingly be used.

The following are the equivalents of the derivative scores used, in terms of tenths of the original inter-group interval scores:

1	67-75	4	36-47	7	4-13
2	58-66	5	25-35	8	-5- 3
3	48-57	6	14-24	9	-26--6

The final score obtained by averaging the (derivative) scores of the various items for a given subject is also best expressed in tenths, i.e., an average of 6.19 is to be expressed as 62; the norms (see "Norms") are tabulated on this basis.

DEGREE RATINGS

The distinction between ratings of 3 and 2 (observed and estimated presence of the trait respectively) is without value and is abandoned in the final scale, since

1. The number of 2's is inadequate. They are nearly absent among the self-ratings (see Self-ratings); among the other ratings the distribution is:

		0-	5-	10	15	20	25	30	35	40-	45-	Total
Rater 1	2	21	23	11	5	6	4					70
	3	7	2	12	10	15	19	2	2	1		70
	2+3				6	11	22	21	7	2	1	70
Rater 2	2	17	19	18	9	3	4					70
	3	3	8	14	17	15	8	3	2			70
	2+3			3	2	14	23	20	4	4		70

2. There are no significant differences between the means and variabilities of the 2's and 3's:

	Rater 1		Rater 2	
	2	3	2	3
Mean	50.6	48.9	49.9	50.7
S.D.	13.1	12.9	12.6	12.0

The greatest difference between the means is .84 and that between the standard deviations is .74 times the respective standard errors.

3. The correlations between scores derived from 2's and 3's are slight: Rater 1, $.14 \pm .08$; Rater 2, $.16 \pm .08$.

4. The reliability of rating of scores based on 3's alone is much less than that based on 2's and 3's together, viz., $.15 \pm .08$ as against $.54 \pm .06$.

Dr. P. E. Vernon, after trying out the original scale on 25 freshmen studied intensively, believes that a score consisting of the average scale value for the 2's and 3's minus that for the 0's and 1's is of merit. In particular, "A comparison of this sort between the 2-3's and the 0-1's turns out to be an excellent measure of what I

call impulsiveness," presumably, however, in the rater rather than in the subject.

RELIABILITY

The term reliability here refers principally to the amount of agreement between two raters of the same subject. For the raters and subjects used (college men and women) this is in the near vicinity of .54 for the final scale (the slight variations depending on the form of scoring device used, etc., may be found under "Scoring" and "Degree Ratings"); it is probable, on the basis of a few exploratory attempts with better trained raters well acquainted with their subjects, that for such persons it can be considerably increased, perhaps to .70 or .80.

In the section on "Consistency of Judgment" may be found the results of an individual experiment on the reliability of *judgment* by one of the judges.

Dr. P. E. Vernon, rating 25 freshman men with whom he had worked intensively, reports: "The rank correlations are: self with P. E. V. .54, self with outside rater, .23, outside rater with P. E. V. .37. That is to say, the corrected reliability of the final, combined order, is .65." Dr. Vernon was here using the original scale (150 items), a tentative set of values, and a scoring system consisting of the average of the 3's+2's minus the average of the 1's+0's. Clinically, he feels that "the final orders do seem to represent something in the individuals, in spite of the difficulties."

In one case a subject (male) was rated by two men and two women besides himself. There was complete agreement between the four raters on 29 of the 60 items of the final scale (15 positive and 14 negative), the subject also agreeing in 11 positive and 12 negative. There was majority agreement, in addition, on 14 positive items and 10 negative, the subject also agreeing with the majority in 9 positive and 9 negative. In 7 items the item was evenly split between positive and negative, with the subject himself estimating the item positively in 3. In terms of score this is equivalent to 61 without the disputed items, 58 with 3 of the 7, and 55 with all of them.

In an experiment conducted some months after the preliminary investigations here reported, by Dr. O. L. Harvey of the University of Texas, who used a form of the final scale with simplified wording, the reliability in the usual sense of correlation between test and re-test (six-weeks' interval) was about .71 (123 cases). In this con-

nection, it is to be noted that Dr. H. F. Adams (1, p. 126, footnote) finds that his "accuracy correlations" (correlations of estimates with objective values) are the square roots of the corresponding "group consistency correlations" (correlations between estimates). The square root of .71 is only slightly lower than .54, which may indicate that the scale, when used for self-rating (as in Harvey's experiment) may yield a fair approximation to a real entity (see also "Validity"). Adams' finding appears, furthermore, to be identical with Kelley's formula 160 (2, p. 206) for the correlation between one form of a test and a true score on the function measured by the test:

$$r_{1\infty} = \sqrt{r_{11}}$$

VALIDITY

Little can be said on the subject of validity; the logic of the method of derivation leaves the question largely without meaning. In the terminology of statistics, no criterion is available. That is, it has been sought to measure a variable for which there is no simple objective correlate; ergo, when the subjective judgments of a large number of experts have been pooled, we have as valid a scale as is possible in the present state of knowledge. With respect to the determination of the presence of a given trait, the matter is one of reliability of rating, treated fully above.

A result derived by Dr. P. E. Vernon from his 25 Yale freshmen bears upon the question of validity from a slightly different angle. Using the original scale of 150 items and a tentative (reversed) scoring scheme, he found "that this final score turns out to be my best measure of emotional instability out of about 20 available ones. Its highest correlations are .60 with an adaptation of Woodworth P. D., and .57 with pathological Kent-Rosanoff responses."

Dr. O. L. Harvey, in two self-rating experiments on college students, using a form of the final scale simplified only in terminology, found in the first case (45 cases) a correlation with the Thurstone Personality Schedule of .43, and in the second (128 cases) one of .10, but with η 's of .24 and .33 (respectively η_{wt} and η_{tw}). With an unpublished extroversion-introversion scale the correlation was .16 (45 cases).

SELF-RATINGS

The scale may be used as a test without large loss of accuracy,

since the usual tendency to rate oneself higher than others rate one does not appear, and the self-ratings correlate nearly as well with the ratings by others as the latter do with each other. The latter coefficients are:

	Self	Rater 2
Rater 1	.49±.06	
Rater 2	.38±.07	.54±.06

Of 70 individuals, each rated twice, 34 rated themselves higher than they were rated by "Rater 1"² and 30 lower; and 29 rated themselves higher than they were rated by "Rater 2"² and 33 lower.

The mean of the self ratings is the same as the other means, viz., 51.2. The standard deviation is a little less, viz., 6.23; this increased certainty is corroborated by the high proportion of 3's among the total positive ratings, viz., of the aforesaid 70 individuals, 46 had all positive self-ratings 3's, and the number of 2's were distributed thus:

1	2	3	4	5	6	..	8	..	10	..	15	Total
10	3	3	1	1	3	..	1	..	1	..	1	24

SEX DIFFERENCES

There are no significant sex differences in the central tendencies of the scores. The means of the 2+3 scores for "Rater 1" and "Rater 2" for 35 unselected persons of each sex are:

Women		Men	
Rater 1	Rater 2	Rater 1	Rater 2
51.4	51.9	50.9	51.7

The greatest difference is .6 of its standard error.

There is, however, a tendency for the women's variability to be greater; since the women were rated almost exclusively by women and the men by men, nothing in the data can be used to determine whether the source of this tendency resides in the subjects or in the raters. The standard deviations corresponding to the above means are:

²Composite "individuals," it should be remembered.

Women		Men	
Rater 1	Rater 2	Rater 1	Rater 2
8.45	7.48	6.30	6.37

The greatest sex difference is 1.7 and the least 1.2 times its standard error; these correspond to odds of about 19 and 5.5 to 1, respectively.

Since practically all the subjects were college students, nothing can be said of age differences.

NORMS

On the basis of the self-ratings and two other ratings of 70 college students, equally divided as to sex, the following smoothed percentile norms are presented:

Score (average x 10)	Percentile
35	1
36	1.5
37	2
38	2.5
39	4
40	6
41	9
42	12
43	16
44	20
45	27
46	30
47	37
48	43
49	48
50	53
51	59
52	64
53	68
54	72
55	76
56	80
57	84
58	87
59	89
60	92
61	94
62	95
63	96
64	97
65	98
66	99
67	99.5

The median is 49.3, and the distribution therefore a trifle asymmetric (mean 51.4); Q_1 and Q_3 are 44.8 and 54.7, respectively.

THE SCALE

The complete scale is given below, with the items of the final selected scale starred; the five figures are, in order, the original scale value (in tenths of an inter-group interval), the derivative value to be used with the selected scale, the spread (Q of judgments) in tenths of an interval, the percentage of disagreement, and the incidence.

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
*1. <i>S</i> is ordinarily friendly toward members of his immediate social group, but in critical periods becomes irritable and hostile.	52	3	10	32	31
2. <i>S</i> lavishes considerable overt affection upon younger person—petting, using as a toy, etc.	70		11	25	9
3. <i>S</i> is friendly and companionable toward his immediate family, without striking emotion.	9		9	21	94
4. <i>S</i> rejects with some emotion any attempt to impose a suggestion upon him.	66		11	33	9
5. <i>S</i> 's efforts in the pursuit of an object are usually persistent and sustained.	11		6	24	95
*6. <i>S</i> is extremely solicitous of his immediate family associates.	51	3	15	37	63
*7. <i>S</i> makes his plans with objective reference to his own death when this issue is involved; no emotional reaction involved greater than that, for instance, concerned in planning with reference to a long journey.	4	7	8	41	49
8. <i>S</i> refuses to take action on basis of folk-lore or superstition, but with evidence of some conflict, e.g., acknowledgment of validity of belief in question, more opposition than situation warrants, etc.	44		9	44	37
*9. <i>S</i> is meticulous in matters of dress; a considerable part of his income is spent in this activity, even though strict economies are thereby necessitated elsewhere.	53	3	8	31	28
*10. <i>S</i> chooses courses of action with reference to his own maximum immediate satisfaction.	75	1	15	34	39

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
*11. <i>S</i> develops affective difficulty in the presence of a necessity for precise or realistic thinking, e.g., mathematics.	58	2	10	30	24
*12. <i>S</i> is faced with an instance of violation of his mores; he is intellectually interested, without emotional shock, and seeks to discover what motives and satisfaction are involved from the standpoint of the violator.	0	8	5	36	69
13. <i>S</i> evinces hostility toward his immediate family associates.	78		6	8	1
14. <i>S</i> reacts to the presence of evil in the world by fantasy, in which the fact is denied.	81		5	13	1
*15. <i>S</i> is deprived of a much anticipated opportunity; he redoubles his efforts to gain just this objective.	27	5	13	34	70
16. <i>S</i> 's efforts in the pursuit of an objective are not persistent or sustained.	59		13	27	8
*17. <i>S</i> characteristically appeals for help in the solution of his problems.	68	1	12	29	20
18. <i>S</i> tends to solve his problems by rational adjustment to actual conditions, by using his intelligence to reach the objectives he desires.	2		5	19	95
19. <i>S</i> is faced with decision in which his own convictions are the reverse of those held by his social group; he decides publicly as he is expected to decide, but secretly takes steps to attain his own ends.	49		12	33	18
20. <i>S</i> is subjected to public criticism, e.g., a reprimand from some person in authority or open adverse criticism of his actions. <i>S</i> becomes incoherent, the victim of a general emotional outburst dominated by anger but without much intellectual content (e.g., some types of weeping in women).	83		3	16	4
21. Manner is unostentatious without being retiring.	6		4	31	88
22. <i>S</i> persistently contemplates suicide as a solution for his problems.	89		10	6	1
*23. <i>S</i> is rather self-conscious in the presence of individuals of markedly greater prestige than his own.	43	4	8	36	34

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
24. <i>S</i> is gossipy in conversations; tacitly solicits the co-operation of interlocutors in remarks about others not present.	64		11	15	8
*25. <i>S</i> conducts himself in discussion as if the only objective of the discussion were the mutual discovery of truth.	7	7	4	41	59
*26. When driving an automobile, <i>S</i> is unperturbed in ordinary situations, but becomes angry with other drivers who impede his progress.	50	3	10	41	36
27. <i>S</i> reacts to presence of markedly younger person by inaugurating a competition for favor or advantage.	68		9	22	4
28. <i>S</i> indulges rather consistently in alcohol or drugs.	81		10	7	1
*29. <i>S</i> 's day-dreams represent the reversal of situations humiliating in the real world.	69	1	11	34	21
30. <i>S</i> 's manner is distinctly of the retiring or withdrawing sort.	63		10	17	11
*31. <i>S</i> believes in democracy in principle, but prefers not to associate too closely with individuals from groups widely divergent from his own.	38	4	7	43	74
32. <i>S</i> usually accepts unwelcome invitations, feeling that nothing else can be done; but is inclined to work off his annoyance in fault-finding, etc.	61		12	25	11
33. <i>S</i> gives attention only to the attainment of his ends in the easiest and most direct manner, neglecting system, precision, etc., in the manner of their attainment.	51		15	34	20
*34. <i>S</i> demands that he be punctiliously served in hotels, sleeping cars, etc.	58	2	12	31	23
35. <i>S</i> is bored and uninterested in conversation, appears to disdain to mingle with the group.	67		10	17	1
36. <i>S</i> is deeply injured at the existence of sexual phenomena, but endeavors to accept facts as they are and make the best of them; reacts to sexual facts as to facts of vice or illness.	59		13	24	7
*37. <i>S</i> passes rapidly from one interest or attachment to another.	60	2	10	31	19
38. <i>S</i> cannot bring himself to harbor thoughts of his own death; struck with emotional panic at any reference to it.	80		6	10	1

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
*39. <i>S</i> organizes and orders his efforts in pursuing his objectives, evidently regarding systematic method as a means of achieving them.	8	7	8	28	82
40. <i>S</i> is subjected to public criticism, e.g., a reprimand from some person in authority or open adverse criticism of his actions. <i>S</i> is silent and evidently surly and resentful; possibly forms elaborate plans for retaliation.	71		9	25	7
41. <i>S</i> takes serious action on basis of folk-lore or superstition (defined as current belief unverified and inherently improbable).	79		8	10	1
42. With younger person, <i>S</i> tends to withdraw from active participation in the group activity, and to observe younger person.	46		10	26	8
43. Upon threatened danger to person affectively near <i>S</i> , <i>S</i> loses self-control—throws up hands, implores aid, etc.	80		4	14	6
*44. <i>S</i> dislikes the labor involved in precise thinking, but when necessary performs it without remarkable emotional disturbance.	30	5	9	43	56
45. <i>S</i> relates at some length and with some satisfaction accounts of the repelling of sexual aggression.	73		16	16	6
46. <i>S</i> relies greatly upon supernatural forces to take decisive part in the solution of his problems.	76		10	14	5
47. <i>S</i> is co-operative in conversation—contributes items of somewhat sublimated and general interest, then waits for the expression of others; is at ease and serene.	7		5	30	89
*48. <i>S</i> can be easily persuaded to comply with the will of a person having some natural authoritative relation to him (e.g., a parent) but is relatively impermeable to suggestions from others.	46	4	11	42	38
*49. <i>S</i> takes a rationalistic viewpoint of evil; desires to learn more about the problem and to put into practice that which he has learned.	2	8	7	27	84
50. <i>S</i> tends to meet his problems by collapse and submission.	83		6	12	1
*51. <i>S</i> is fair and courteous in games, but feels that much depends on the outcome.	38	4	8	37	68
*52. <i>S</i> is usually on terms of equality with his immediate social group, but in critical periods is submissive to or dominated by them.	45	4	7	31	32

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
53. <i>S</i> assumes the satisfaction of his desires to be a natural right.	65		25	33	52
*54. Faced with several insistent demands simultaneously, <i>S</i> is upset, and fights the situation; his only idea is to compel order.	63	2	10	35	15
55. <i>S</i> reacts to thwarting by combativeness and increased incentive to dominate.	59		16	36	38
56. <i>S</i> is relatively constant in the degree and duration of his attachments and interests.	11		7	23	91
57. <i>S</i> is faced with an instance of violation of his mores; he is emotionally crushed, feeling that the world can never be the same again, etc. (especially if he has a personal relation to the violator).	79		7	19	5
58. <i>S</i> 's day-dreams represent him striving with and overcoming difficulties.	46		18	43	49
59. <i>S</i> attaches considerable importance to the satisfaction of his desires, but realizes that they must frequently remain unsatisfied.	18		8	23	100
*60. <i>S</i> is "analytic" in conversation; inhibits own expression for therapeutic ends.	22	6	12	39	25
*61. <i>S</i> is exposed to unsocial behavior (e.g., antagonism) on the part of an individual with whom he is closely associated; he is annoyed, and seeks to escape or to drive the other individual away.	61	2	11	37	17
62. <i>S</i> is deprived of a much anticipated opportunity; he flies into a rage ("temper tantrum").	84		6	12	5
*63. <i>S</i> tends to take on the psychical characteristics of admired individuals.	51	3	12	38	18
64. <i>S</i> drives an automobile with much speed and daring, deriving evident exhilaration from the process.	52		13	31	23
65. <i>S</i> is exposed to unsocial behavior (e.g., antagonism) on the part of an individual with whom he is closely associated; he adopts an attitude of amused superiority.	45		11	36	20
*66. <i>S</i> chooses a course of action with reference to maximum long-time satisfaction of entire group affected.	4	8	4	33	70
67. <i>S</i> refuses to accept a suggestion; he does not manifest any emotional disturbance in connection with the refusal, but also gives no evidence of having considered the suggestion.	48		9	42	19

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
*68. <i>S</i> is "scientific" in conversation; appears to regard others' viewpoints as specimens.	37	4	13	35	25
69. <i>S</i> feels more comfortable when his professional and personal affairs are kept separate than otherwise.	34		13	45	56
70. <i>S</i> accepts service easily and naturally, being pleasant and unostentatious to servants, not demanding more than needed or leaving openings for insubordination.	-4		6	20	92
71. <i>S</i> is troubled by night terrors or terror dreams.	76		6	10	4
72. Upon introduction into group of distinctly younger persons, <i>S</i> reacts in patronizing or condescending way, carrying implication that his courtesy is conditioned upon clear recognition of his superiority.	65		11	26	7
73. <i>S</i> devotes much energy to recounting his achievements.	61		10	26	14
*74. <i>S</i> can react without devastating emotion to the thought of his own death, but only with the help of some negating belief, e.g., survival.	44	4	7	33	24
75. <i>S</i> is deprived of a much anticipated opportunity; he is emotionally crushed and incapable for a time of further effort in any direction.	80		6	24	8
*76. <i>S</i> chooses course of action in accordance with estimated maximum long-time satisfaction for himself.	37	4	13	36	73
77. <i>S</i> takes offence at negative reaction to him on part of distinctly younger person with whom he is thrown, e.g., crying of child, shrinking, fleeing, dislike, etc.	70		9	21	8
78. <i>S</i> is submissive in conversation; seems to be interested only in the attainment of others' satisfactions.	48		13	24	11
79. <i>S</i> is prying in conversation, curious about the affairs of interlocutors, "pumping."	60		11	22	6
80. <i>S</i> finds emotional release from his difficulties in sex gratification.	68		17	17	6
*81. <i>S</i> unostentatiously admits younger person into group of which he is a member, continues to take part in activities of expanded group. No difference evident between <i>S</i> 's conduct toward younger person and any other.	7	7	5	30	83

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
82. <i>S</i> resents existence of sexual phenomena as personal affront to be warded off. (E.g., "All men are brutes," offered by a woman.)	79		6	13	1
83. <i>S</i> is characteristically apprehensive concerning the possibility of sexual aggression.	75		10	28	8
84. <i>S</i> conducts himself in discussion as if the object of the discussion were demonstration of his personal superiority.	66		12	22	5
85. <i>S</i> is obviously in awe of servants, etc., so that he finds difficulty in asking for what he desires or goes without it. Manner an invitation to insubordination.	73		17	16	1
*86. <i>S</i> is faced with an instance of violation of his mores; altho emotionally jarred, he seeks intellectually for a course of action consonant with his self-ideal and endeavors to carry it out.	21	6	9	35	82
*87. <i>S</i> is "jealous" of his spouse, feels insecure when any other interest claims spouse's attention.	72	1	11	36	28
88. <i>S</i> is deeply shaken by pain, giving the impression that a psychic as well as a physical factor is involved.	68		11	25	12
89. <i>S</i> is deeply shocked by the presence of evil in the world; is crushed, and resigned to the inscrutable ways of Providence.	74		11	9	1
90. <i>S</i> tends to meet his problems by ascribing elsewhere the responsibility for them.	73		15	23	12
91. <i>S</i> is angry at his inability to attain his objectives, and blames his tools or environment for the failure.	75		7	23	13
*92. Being interrupted in the performance of an act, <i>S</i> returns to it at the earliest opportunity, giving evidence of discomfort at the diversion.	38	4	9	35	40
93. <i>S</i> withdraws in a conversational situation; fears to inject his point of view unless invited.	55		8	24	9
*94. <i>S</i> is much attached to his regional or social group, seizing every opportunity to extol its merits.	50	3	10	34	48
95. <i>S</i> is subjected to public criticism, e.g., a reprimand from some person in authority or open adverse criticism of his actions. <i>S</i> collapses, shows signs of emotion (e.g., weeping) and submission without defense.	80		4	8	2

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
96. <i>S</i> secures much satisfaction from the attempted perpetuation of previous experiences; e.g., he retains strong identifications with his college or military associations.	47		8	43	51
97. <i>S</i> is deprived of a much anticipated opportunity; he readjusts his course of action in such a way as to orient toward a revised objective.	6		4	21	92
*98. <i>S</i> does not evince interest in why an action should be performed, feeling very strongly only that it should or should not.	56	3	11	33	22
99. <i>S</i> takes jocular or apologetic action on basis of folk-lore or superstition (defined as current belief unverified and inherently improbable).	51		11	40	22
*100. <i>S</i> is scrupulously tidy, placing neatness high among his major objectives.	45	4	10	34	56
101. <i>S</i> 's manner is unnecessarily bluff, blustering, or if quiet, insistent and pertinacious. Net effect is to keep <i>S</i> in the foreground of attention a large part of the time.	66		11	23	9
102. <i>S</i> 's conversation is a monologue—he seems unable to avoid doing all the talking.	66		10	14	6
*103. <i>S</i> feels that the universe is relatively impersonal and non-ethical regarding his wishes and those of his group or of mankind.	9	7	15	40	37
*104. <i>S</i> can be persuaded rather easily to comply with the will of another, even when the other has no necessary hold upon him.	66	2	12	30	17
105. <i>S</i> is sometimes objective and sometimes affectively motivated in discussions.	39		7	31	87
*106. <i>S</i> dramatizes narrations freely, demonstrates exact manner in which everything happened.	44	4	9	39	36
*107. <i>S</i> evaluates suggestions emanating from any person without heat, settling the issue upon rationalistic bases; and cannot be persuaded to alter a matured decision except on the basis of a new evidence.	-13	9	15	31	78
108. <i>S</i> gains his ends by direct means, rarely trying to divert attention from the issue.	10		7	28	93
109. <i>S</i> carries favor of younger persons with whom he is thrown by offers of gifts, extraordinarily attentive behavior, etc.	71		11	18	6

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
110. <i>S</i> is deprived of a much anticipated opportunity; he "resigns" himself to the change—i.e., is crushed, impressed with the fleeting character of human satisfaction, etc.	68		11	22	6
111. Upon introduction into group of younger persons, <i>S</i> tends to react by domination or appeal to competition, i.e., seeks to establish an order of precedence or superiority.	63		10	21	6
112. <i>S</i> tends to form strong attachments, sometimes to distant individuals; "hero-worshipping."	54		13	22	21
113. <i>S</i> disregards younger persons with whom he is thrown, and proceeds about his own concerns.	44		12	38	23
*114. Being interrupted in the performance of an act, <i>S</i> gives attention, when resumption is possible, to the total situation, and resumes or abandons the act unemotionally as the new situation makes advisable.	-3	8	8	24	83
*115. <i>S</i> is exposed to unsocial behavior (antagonism, etc.) on the part of an individual with whom he is closely associated; he adopts a detached attitude and is interested in the underlying causes.	3	8	7	33	55
116. <i>S</i> reacts fully to pain, but does not dread it unduly or continue reacting to it when no longer present.	4		10	27	95
117. <i>S</i> definitely disregards (without emotional disturbance) implication that he should take action on basis of folk-lore or superstition (defined as current belief unverified and inherently improbable).	5		7	31	87
*118. <i>S</i> is strongly conscience-ridden; anxious lest he violate the sanctioned codes.	64	2	10	30	21
*119. On unmistakable demonstration of his inferiority in some respect, <i>S</i> is impressed, but consoles himself by the contemplation of those activities in which he is superior.	32	5	11	40	72
120. <i>S</i> handles realistically and effectively the ordinary problems of his existence, but is upset by critical ones and behaves emotionally or irrationally.	51		10	29	24
121. <i>S</i> tends to meet his problems by attempts to day-dream them away.	79		9	18	5

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
*122. <i>S</i> derives evident satisfaction from bringing to attention instances of his own usefulness, altruism, etc.	57	3	12	30	16
*123. Faced with several insistent demands simultaneously, <i>S</i> sorts them quickly by urgency, settles one at a time, disregards (i.e., does not react to) insistence of deferred demands.	3	8	8	30	73
*124. <i>S</i> is "resigned" to thought of his own death—i.e., has accepted its inevitability, but feels it as an enormous calamity.	47	4	10	28	17
*125. <i>S</i> welcomes opportunity for exercise of precise or realistic thinking, e.g., mathematics.	9	7	10	33	43
126. <i>S</i> tends to react to his problems by avoiding them.	74		10	19	4
127. Upon threatened danger to person affectively near <i>S</i> , <i>S</i> seriously and calmly takes steps designed to secure best issue.	-8		8	24	91
*128. <i>S</i> feels that the universe (God, etc.) takes an active interest in his affairs and those of his group of mankind.	52	3	14	31	51
129. <i>S</i> evaluates large share of his experiences in terms of the opportunity they afford for sexual gratification.	76		16	15	7
130. <i>S</i> is impulsively eager to alleviate evil; does not see any need for deliberation, which appears to him as temporizing. Impatient with "theories," wants to go and fight it.	60		11	25	11
*131. <i>S</i> urges child or younger person to "perform," i.e., exhibit for group approval any abilities he may possess.	51	3	10	33	36
*132. <i>S</i> disregards all but realistic, naturalistic factors in the solution of his problems.	14	6	15	42	47
133. Being interrupted in the performance of an act, <i>S</i> has forgotten about it when opportunity again permits its resumption.	62		13	28	6
134. <i>S</i> is subjected to public criticism, e.g., a reprimand from some person in authority or open adverse criticism of his actions. <i>S</i> becomes angry, resorts to <i>ad hominem</i> tactics and reprisals.	75		10	19	5
135. <i>S</i> reacts to attempts to solve his problems as to a threat, viz., by defensive tactics.	63		12	32	12

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
*136. <i>S</i> welcomes legitimate opportunities for sexual expression, is not ashamed, fearful or preoccupied with the topic.	-26	9	9	32	78
*137. <i>S</i> is exposed to unsocial behavior (e.g., antagonism) on the part of an individual with whom he is closely associated; he reacts by attempted domination—seeking to overcome the resistance implied.	49	3	9	39	23
138. Faced with several insistent demands simultaneously, <i>S</i> collapses, calls for help.	84		5	14	5
139. <i>S</i> feels that books, plays, etc., should be clean and uplifting, since life brings so much that is inevitably otherwise anyway.	42		18	31	51
*140. <i>S</i> is subjected to public criticism, e.g., a reprimand from some person in authority or open adverse criticism of his actions. <i>S</i> waits composedly for opportunity to reply, states his position without heat, possibly then departs.	5	7	5	33	70
141. <i>S</i> is preoccupied in conversation; his attention is evidently diverted by other interests.	53		10	24	7
*142. <i>S</i> is tolerant of most divergent opinions, but feels there are some ideas so far beyond the pale as to make their extinction an obligation on all right-minded people.	42	4	14	34	79
*143. <i>S</i> is accustomed to achieve some of his ends by indirection, e.g., by diverting the attention of important individuals to other interests.	43	4	10	35	17
*144. <i>S</i> is serious and anxious in his manner, even in cases where nothing important can hang on the results; "worried."	65	2	10	33	21
*145. <i>S</i> is clear-cut in his decisions; when relinquishing an objective he relinquishes or postpones it entirely, when retaining it he retains all of it without regret.	9	7	6	36	67
146. Faced with several insistent demands simultaneously, <i>S</i> is calm or possibly good-humored, but evades the decision, perhaps by delegating it to someone else.	44		8	34	15
147. <i>S</i> is faced with an instance of violation of his mores; he reacts jocularly, makes the violator the butt of jokes, etc.	51		9	25	11

Item	Orig. value	Der. value	Spread	Disagreement	Incidence
148. Faced with several insistent demands simultaneously, <i>S</i> is calm, but does nothing to settle the demands until he has first restored order.	18		13	48	67
149. <i>S</i> is the captain of his soul with respect to his own death; he has resolved to go down defying the gods, etc.	52		14	29	17
*150. <i>S</i> cannot give up or retain an objective completely; if he tries to give it up, he must make some gesture of retention, and if he retains it he feels the desirability of giving it up.	60	2	15	30	30

SUMMARY

A scale of 60 items has been devised for the estimation by a rater of the degree of emotional maturity of a subject. This is effected by checking those items which describe characteristic attitudes and reactions of the subject, and scoring the checked items by the use of values derived from the pooled judgments of 100 personality experts.

The reliability of rating with the instrument is about .55 for college student raters, and is independent of the technique employed for combining the item values; it is probably higher for better trained raters. The test-retest reliability is about .70.

There will probably be a medium to high correlation with results from instruments alleged to measure emotional stability.

Both actual and probable behavior must be taken into account in the estimate.

The usual tendency to over-rate oneself is not found in self ratings with this scale; the scale may therefore be used as a test with only a small loss of precision.

There are no sex differences in central tendency, but women raters rating women subjects show a slight tendency to increased variability.

About 75 items are presented in addition to the 150 from which the selection of 60 was made; given adequate time and resources, these can be standardized and used to repair deficiencies in the present

scale. A rich field is also open for the discussion and analysis of the fundamental concepts, which are felt to be of great practical and theoretical importance.

REFERENCES

1. ADAMS, H. F. An objectivity-subjectivity ratio for scales of measurement. *J. Soc. Psychol.*, 1930, **1**, 122-135.
2. KELLEY, T. L. Statistical method. New York: Macmillan, 1924. Pp. xii+390.

Clark University

Worcester, Massachusetts

UNE ÉCHELLE DE MATURITÉ ÉMOTIVE

(Résumé)

On a fait une échelle de 60 parties pour l'estimation par un juge du degré de maturité émotive d'un sujet. La méthode d'obtenir celle-ci a été la suivante: on a marqué les parties qui décrivent les attitudes et les réactions caractéristiques du sujet, et on a évalué les parties marquées par l'emploi des valeurs obtenues de l'ensemble des jugements de 100 experts en personnalité.

La valeur de l'évaluation avec l'appareil est d'environ 0,55 pour les étudiants universitaires servant de juges, et ne dépend pas de la technique employée pour combiner les valeurs des parties. Elle serait plus élevée pour les juges mieux entraînés.

Il y aura une corrélation moyenne ou élevée avec les résultats des appareils censés de mesurer la stabilité émotive.

On doit tenir compte du comportement vrai et du comportement probable dans l'évaluation.

On ne trouve pas la tendance usuelle à s'évaluer trop dans les évaluations de soi avec cette échelle; on peut donc employer l'échelle avec une très petite perte de précision seulement.

Il n'y a pas de différences de sexe dans la tendance centrale, mais les juges femelles qui évaluent des sujets femelles montrent une petite tendance à plus de variabilité.

On présente environ 75 parties outre les 150 dont on a fait le choix de 60; si l'on a assez de temps et de ressources, on peut les standardiser et les employer pour améliorer l'échelle actuelle. Il se présente aussi beaucoup d'occasions pour la discussion et l'analyse des concepts fondamentaux, qu'on considère d'une importance extraordinaire pratique et théorique.

WILLOUGHBY

EINE STUFENSERIE ZUR MESSUNG DER AFFEKTIVEN REIFE

(Referat)

Der Verfasser hat eine aus 60 Punkten bestehende Phrasenserie (scale) erfunden, die einem Versuchsleiter zur Abschätzung des Grades der affektiven Reife (emotional maturity) einer Versuchsperson dienen soll. Dies wird getan indem man diejenigen Punkte aufzählt, die charakteristische Einstellungen und Reaktionen der Vp. beschreiben, und dann die aufgezählten Punkte zensiert (score) mittels Gebrauch von Werten, die durch die totalisierten Urteilen von 100 Persönlichkeitskundigen (personality experts) erhalten worden sind.

Die Zuverlässigkeitsziffer (reliability of rating) der Zensierung mittels der Skala (instrument) ist, wenn Studenten die Zensierungen machen, ungefähr .55. Es kommt dabei nicht darauf an, auf welche Weise man die Punktwerte (item values) kombiniert. Bei besser eingeübten Zensuristen (raters) wird die Zuverlässigkeitsziffer wahrscheinlich höher sein.

Es wird wahrscheinlich eine mittelmässige bis hohe Korrelation bestehen mit Befunden, die an Werkzeugen erhalten worden sind welche die affektive Festigkeit (stability) messen sollen.

Sowohl das eigentliche wie das wahrscheinliche Benehmen (behavior) muss in der Schätzung in Betracht gezogen werden.

Die gewöhnliche Neigung, sich zu hoch abzuschätzen, macht sich bei Selbstschätzungen mittels dieses Massmittels nicht bemerkbar. Die Skala kann deshalb als Prüfung (test) gebraucht werden, ohne viel dabei an Genauigkeit zu verlieren.

Es zeigen sich keine Geschlechtsunterschiede in der zentralen Richtung (central tendency) [der Befunde]. Weibliche Zensuristen erweisen aber bei der Zensierung weiblicher Vpp. eine Neigung nach gesteigerter Variabilität.

Es werden ausserhalb der 150 Punkte, worunter die 60 auserlesen worden sind, auch 75 weitere Punkte dargestellt. Bei hinreichender Zeit und hinreichenden Mitteln können auch diese standardisiert werden und zur Verbesserung von Mängeln in dem gegenwärtigen Massmittel gebraucht werden. Es eröffnet sich auch ein reiches Feld zur Besprechung und Analyse der grundlegenden Begriffe, die sich als praktisch und theoretisch von grösster Wichtigkeit empfinden lassen.

WILLOUGHBY

A STUDY OF THE QUESTIONS OF YOUNG CHILDREN CONCERNING SEX: A PHASE OF AN EXPERIMENTAL APPROACH TO PARENT EDUCATION*¹

From the Iowa Child Welfare Research Station

KATHARINE WOOD HATTENDORF

This study was undertaken to clarify in the minds of parents the sex information being sought by children in their homes in the early years, as shown by questions gathered from a large number of parents whose children were under junior-high-school age at the time the interviews with the mothers were held. The specific questions that this study aimed to answer were the following: (a) Do young children, in general, ask questions concerning sex? (b) If so, at what specific ages are questions likely to come? (c) What items of information do children seek?

Opportunity for the accumulation of questions of children concerning sex was made possible through the Women's Co-operative Alliance, a community organization for the promotion of social hygiene in Minneapolis, Minnesota. The program of sex education for parents has been developed in the Department of Parent Education, one of the four departments of the same organization. The questions were obtained in the routine home-calling during a period of 17 months, excluding vacations. They form a part of the regular information which is being secured from the experience of mothers by means of home contacts. The home-calling done within this period was limited to the West High School district which includes 17 grade schools. This district had been chosen for the research project being conducted by a joint committee from the Women's Co-operative Alliance and the University of Minnesota with funds contributed by the Bureau of Social Hygiene, New York City. Its evident economic stability and the educational advantages of the group represented made it suitable for this purpose.

*Recommended by George D. Stoddard, accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office, December 23, 1930.

¹A report of a Master's thesis, 1929, directed by Mrs. May Pardee Youtz, and carried out at the Iowa Child Welfare Research Station.

The staff included the director; the writer, as supervisor of field work; six parent-advisers, including a physician, a graduate nurse, a psychiatric social worker, and two social workers, all of whom had had previous teaching experience; a secretarial assistant; and a secretary of records.²

The morning hours of the five parent-advisers were spent in making home calls, and home interviews were confined to mothers with children of preschool and early grade-school age. Homes with children over 14 years of age were excluded, as mothers of adolescent boys and girls were considered less likely to utilize the help of the organization if they had not previously discussed the subject of sex with their children and had not gained their children's confidence.

The term "interview," as applied to the home call, designates any conversation with a mother in which she makes a definite statement in reference to the presentation of sex education to children. Complete interviews during the time this study was made were from 40 minutes to 1 hour in length. The technique used in approaching the mother at the door was a simple one. The parent adviser announced herself as "Mrs. A . . . of the Women's Co-operative Alliance." If the organization was known to the mother, she was in most instances responsive to the call of the parent adviser, unless extremely busy. Information was then gained as to whether there were children of the prescribed age in the home. The parent adviser stated clearly early in the conversation that the organization is interested in knowing what mothers are doing in reference to sex education for their children, and the mother's attitude was thus called forth. Inquiry as to whether the children had already asked questions was found to be a direct method of bringing the mother into the conversation.

If there were inhibitions they were usually expressed by the mother at this time. Many such inhibitions were centered in early experiences which the mother had had and which caused the subsequent reactions. Hesitation on the part of the mother was met by questioning her concerning the sources of her sex information. These are some of the difficult attitudes which the parent advisers had to over-

²Mrs. Robbins Gilman, director, Mrs. William G. King, Dr. Elizabeth Monahan, Miss Julia Pomeroy, Mrs. V. W. Rotnem, Mrs. Frances B. Strain, and Mrs. C. A. Zuppann, parent advisers; Miss D. Fern Smith and Miss Eleanor J. Butler, secretaries.

come in discussing the advisability of giving early sex education to children:

"My child is too young."

"My child is not interested in such things."

"I don't believe a child's mind should be burdened with such things."

"The type of neighborhood in which we live counteracts the need for consideration of this subject."

"My child is too good. She would not listen to sex discussions."

"I believe my own mother's policy of silence was best."

"Won't information stimulate curiosity?"

"I give close supervision to my child. I know just where she is all the time."

"Don't you think sex knowledge just comes intuitively?"

"If the subject is once opened won't the child ask too many questions?"

"Won't information stimulate experimentation?"

"My husband thinks it is unnecessary to give this information."

Where such inhibitions were found mothers naturally had not been receptive to questions concerning sex from their children, their own attitudes on the subject having doubtless kept their children from coming to them openly with questions.

There were 981 interviews among the 3330 recorded for the period in which questions from children were reported. Information from these 981 homes has been used as the basis for this study, inasmuch as the questions reported give evidence that there were children in these homes old enough to question; that the mothers were open-minded in reference to the subject and had recognized interest or lack of interest on the part of their children. There were 1955 children in the 981 families for whom 1763 questions were reported.

A tabulation was first made showing the sex and age distribution of the children (Table 1). In this study, two years of age refers to children who range in chronological age from twenty-four months to thirty-six months. There were 910 boys and 887 girls or a total of 1797 boys and girls between the ages of two and fourteen years, the ages included in the study.

The largest number of children were found in the four-, five-, six-, and seven-year-old groups, there being 192, 209, 203, and 191, respectively, for these ages. The small number for the twelve- and thirteen-year age levels (86 and 52) may be attributed to two

TABLE 1

Years	Boys	Girls	Boys and girls	Percent- age of 1797
2	64	76	140	7.8
3	96	79	175	9.7
4	92	101	192	10.7
5	111	98	209	11.6
6	99	104	203	11.3
7	94	97	191	10.6
8	72	73	145	8.0
9	90	74	164	9.1
10	76	58	134	7.5
11	48	57	105	5.8
12	44	42	86	4.8
13	24	28	52	2.9
Total	910	887	1,729	100.0

causes: first, the calling was done, for the most part, in an outlying district of the city where it was anticipated there would be the greatest number of mothers with young children; and, secondly, the homes in which there were children over fourteen years of age were excluded to avoid the possibility of too great a lapse between the time the questions were asked and the time of recall by the mothers. The records show the ages and sex of the children at the time the home call was made by the parent adviser. They also give the age and sex of each child at the time a question was asked, with questions stated as worded by the child.

ANALYSIS OF THE QUESTIONS OF CHILDREN RELATING TO SEX

In studying the 1763 questions of children obtained in home-calling through the reports of 981 mothers, information was first sought in reference to the ages at which children ask questions concerning sex. A second interest of equal importance was the desire to learn just what children wish to know about sex matters as manifested by their questions. Both of these factors need to be considered in developing a program and materials for early sex education.

The 1797 children between the ages of two and fourteen years, after being classified according to age and sex, were classified according to whether or not they had asked questions. There were 563 boys and 546 girls, making a total of 1109 boys and girls for whom questions were reported. These numbers include 61.9% of the boys and 61.5% of the girls. From these percentages one is led to con-

TABLE 2

CHILDREN FOR WHOM QUESTIONS WERE REPORTED AND FOR WHOM QUESTIONS WERE NOT REPORTED AT THE TIME OF INTERVIEW ACCORDING TO AGE AND SEX

Age, Years	Children for whom questions were reported at one or more ages		Children for whom questions were not reported		Total
	Num- ber	Per cent	Num- ber	Per cent	
Boys					
2	10	15.6	54	84.4	64
3	39	41.7	57	58.3	96
4	49	53.3	43	46.7	92
5	81	73.0	30	27.0	111
6	76	76.8	23	23.2	99
7	68	72.3	26	27.7	94
8	50	69.4	22	30.6	72
9	73	81.1	17	18.9	90
10	45	59.2	31	40.8	76
11	32	66.7	16	33.3	48
12	24	54.4	20	45.5	44
13	16	66.7	8	33.3	24
Total	563		347		910
Mean		61.9		38.1	
Girls					
2	12	15.8	64	84.2	76
3	40	50.6	39	49.4	79
4	47	46.5	54	53.5	101
5	67	68.4	31	31.6	98
6	78	75.0	26	25.0	104
7	65	67.0	32	33.0	97
8	55	75.3	18	24.7	73
9	58	78.4	16	21.6	74
10	46	79.3	12	30.7	58
11	41	71.9	16	28.1	57
12	23	54.7	19	45.3	42
13	14	50.0	14	50.0	28
Total	546		341		887
Mean		61.5		38.5	
Boys and Girls					
Total	1,109		688		1,797
Mean		61.7		38.3	

clude that the degree of interest in the subject as shown by manifested curiosity is equal in boys and girls (Table 2).

The largest number of boys who were reported to have asked questions were six, seven, and nine years old at the time of interview. The largest number of girls who were reported to have asked questions were six, eight, nine, and ten years old at the time of the interview.

This does not mean that their questions were asked at these ages, however. A boy of five, for example, reported as having asked questions, might have asked the question at any of the preceding ages, with no questions occurring at five years. These figures are significant only in relation to the age distribution of children who had or had not asked questions at some period preceding the interview.

A classification was made according to the age at the time each question was asked (Table 3). Here there occurred duplications in the number of boys and the number of girls where questions were asked by the same child at different ages. To meet this difficulty there is included later a study of the first question of each child. In considering the number asking questions at each age, it was necessary

TABLE 3
AGES AT WHICH 563 BOYS AND 546 GIRLS ASKED QUESTIONS AS REPORTED BY MOTHERS

Age, Years	Number at age at time of interview	Number having passed the age at time of interview	Number having passed the age who had asked questions at the age	Per cent
<i>563 Boys</i>				
2	64	910	21	2.3
3	96	846	68	8.0
4	92	750	109	14.5
5	111	658	122	18.5
6	99	547	91	16.6
7	94	448	80	17.8
8	72	354	59	16.6
9	90	282	44	15.6
10	76	192	42	21.8
11	48	116	19	16.4
12	44	68	9	13.2
13	24	24	3	12.5
<i>546 Girls</i>				
2	76	887	28	3.1
3	79	811	75	9.2
4	101	732	92	12.5
5	98	631	124	19.6
6	104	533	84	15.7
7	97	429	69	16.0
8	73	332	58	17.4
9	74	259	29	11.1
10	58	185	40	21.6
11	57	127	24	18.8
12	42	70	7	10.0
13	28	28	6	21.4

to compare the boys and girls who had passed through the age. This comparison shows that the number of boys who asked questions was quite evenly distributed from four to twelve years, with slight peaks occurring at five and at ten years. The percentage who had asked questions varied within this range from 14.5% at four years to 21.8% at ten years. For girls, the number who asked questions was distributed, though less evenly, from the fourth through the thirteenth year with peaks occurring at five, ten, and thirteen years. The percentage varied within this range from 10.0% at twelve years to 21.6% at ten years. The peaks show that 19.6% of the girls at five years, 21.6% at ten years, and 21.4% at thirteen years had asked questions.

Table 4 includes the 1763 questions classified according to the ages at which they were asked. Here similar peaks are found for

TABLE 4
OCCURRENCE OF 1763 QUESTIONS OF 1797 CHILDREN ACCORDING TO AGE AND SEX

Age, years	Number at age at time of interview	Number having passed the age at time of interview	Number at age having asked questions	Number of questions reported for age	Mean questions for child
<i>Boys</i>					
2	64	910	21	29	1.38
3	96	846	68	91	1.34
4	92	750	109	145	1.33
5	111	658	122	176	1.44
6	99	547	91	129	1.42
7	94	448	80	105	1.31
8	72	354	59	81	1.37
9	90	282	44	69	1.57
10	76	192	42	59	1.40
11	48	116	20	25	1.25
12	44	68	9	11	1.22
13	24	24	3	3	1.00
<i>Girls</i>					
2	76	887	28	38	1.36
3	79	811	75	100	1.33
4	101	732	92	111	1.22
5	98	631	124	175	1.41
6	104	533	84	118	1.40
7	97	429	69	91	1.32
8	73	332	58	73	1.26
9	74	259	29	41	1.41
10	58	185	40	49	1.22
11	57	127	24	26	1.08
12	42	70	7	9	1.28
13	28	28	6	9	1.50

both boys and girls, the greatest number of questions for boys having occurred at five (1.44 questions) and at nine years (1.57 questions); for girls, at five (1.41 questions), at nine (1.41 questions), and at thirteen years (1.50 questions).

A study was also made of the occurrence of first questions; these also being considered in relation to the number of boys and girls who had passed through the successive ages and who, therefore, had had opportunity to question at the specified age. The first questions of boys came chiefly between the ages of four and eleven years, with a peak of 16.1% at five years and another peak of 16.7% at ten years. The first questions of girls came chiefly between the ages of four and thirteen years, with peaks of 17.9% at four years, 15.6% at ten years, and 17.8% at thirteen years.

In using for this study questions that had to be recalled by mothers after a varying lapse of time, it is natural for a question to arise as to the reliability of the ages at which children asked questions as reported by mothers. A study of the period which had elapsed between the time 1763 questions were asked and the time they were reported by the mothers shows how recently most of the questions had occurred. Of these 1763 questions, 38.8% had occurred less than one year previous to the report; 21.5% within one year; and 10.6% within two years. There were 239, or 13.6% of the questions for which this lapse could not be definitely estimated due to the fact that there were two children in the same family in these cases of such ages and sex that either might have asked the question at an earlier age. Since so large a number of the questions reported had been recent questions, these 239 questions were counted as having been asked by the child in each family who was nearest the age at which the question was asked. The lapse in time for the total number of questions was then found to be: less than one year, 46.9%; one year, 23.8%; two years, 12.0%; three years, 6.5%; four years, 4.9%; and five to eight years, 5.9%.

Questions were found to be lacking for the early years of the older boys and girls, although children two, three, four, and five years old who were at these ages at the time of the interview were asking many questions at these early years. To overcome this discrepancy, an analysis was made of the questions of boys and the questions of girls who were at the specified ages at which the questions were asked at the time of the interview. Here the factor of memory deflection should be at its minimum. The results show a fairly regular

tendency on the part of both boys and girls to question at all ages from two to ten years for boys and from two to thirteen years for girls. Peaks occur for boys at four years (1.53 questions), at six, (1.57 questions), and at nine (1.48 questions). For girls, the peaks come at six (1.63 questions), at nine (1.59 questions), and at thirteen years (1.50 questions).

Along with the knowledge as to when children seek information concerning sex should go information in reference to what, specifically, children are wanting to know. A study was accordingly made of the subject-matter of the 1763 questions of children which had been submitted by mothers. The questions classified rather definitely into eight groups as shown in Table 5. The questions, thus classified, were checked by two directors of parent education programs, a parent adviser, a mother, a father, and two graduate students majoring in child welfare.

Table 6 shows the distribution by age for each of the groups into

TABLE 5
CLASSIFICATION OF QUESTIONS

Question	Number	Percentage
Origin of babies	722	40.9
Coming of another baby	256	14.5
Intra-uterine growth	42	2.4
Process of birth	183	10.4
Organs and functions of the body	209	11.9
Physical sex differences	226	12.7
Relation of the father to reproduction	92	5.2
Marriage	36	2.0

TABLE 6

RANK OF INTEREST FOR 865 QUESTIONS OF CHILDREN TWO TO FIVE, 707 QUESTIONS OF CHILDREN SIX TO NINE, AND 191 QUESTIONS OF CHILDREN TEN TO THIRTEEN YEARS CLASSIFIED IN EIGHT GROUPS

Classification	Age, years		
	2 to 5	6 to 9	10 to 13
Origin of babies	1	1	2
Coming of another baby	4	2	1
Intra-uterine growth	7	7	8
Process of birth	5	3	5
Organs and functions	3	4	3
Physical sex differences	2	4	6
Relation of father to reproduction	6	6	4
Marriage	8	8	7

which the questions were classified. Through this table, analysis can be made, if desired, of the interests of the children at each single age level. For analysis in this study, the questions were classified into three age levels; questions of preschool children, two to five years; questions of children of early school age, six to nine years; and questions of pre-adolescent children, ten to thirteen years. There were found to be 865 questions or 49.1%, in the first group; 707 questions, or 40.1%, in the second group; and 191, or 10.8%, in the third group.

Comparative interest shown in the various classifications of questions can be shown most clearly for the three age levels, perhaps, through ranking the interest in each group of questions for each age level.

The interests of the preschool group retain the same rank when considered in relation to all the questions asked concerning a single group of questions as was shown when considered in relation to all questions of the preschool group alone. The questions of the early school years, compared with all the questions, emphasize interest in the coming of new babies, the process of birth, and marriage. The pre-adolescent questions show a decline in interests of the preceding years. More knowledge is sought at this time in reference to the father's place in reproduction and the part that marriage plays.

The questions, as listed, represent the types of questions asked, although variations occur in the exact phraseology used by individual children in expressing the questions. The number listed for each group accords with the different ideas expressed.

The interest in the classifications of questions is almost equally divided between boys and girls. Too few cases of successive questions on the part of individual children were reported to give a fair representation of the sequence of interest on the part of individual children. This is, of course, partly due to the fact that relatively few early questions of children have received adequate answers. As a result, many second questions are, in part, repetitions of the first questions asked.

The relative interest shown by questions relating to human beings and those relating to animals can be traced through the eight groups of questions. In all, there are only 214, or 12.2%, of questions relating to animals. This shows clearly that the children's interest is in the realm of human beings who form their environment.

While animals may be valuable as a means of supplementing their knowledge, these questions prove that parents should be prepared to answer questions in the realm of human reproduction and life.

TABLE 7

TWO HUNDRED FORTY REPRESENTATIVE QUESTIONS CLASSIFIED IN EIGHT GROUPS
ACCORDING TO SUBJECT-MATTER FOR CHILDREN TWO TO FIVE,
SIX TO NINE, AND TEN TO THIRTEEN YEARS

Questions	Ages				No. of questions
	2 to 5	6 to 9	10 to 13		
	Boys	441	384	98	
	Girls	424	323	93	840
		865	707	191	1763

Group 1: 722 questions concerning origin of babies

1. Where do babies come from?	Boys	65	50	6	121
	Girls	74	52	11	137
2. Where did the baby come from?	Boys	29	8	1	38
	Girls	17	6	1	24
3. Where did the kittens come from? (puppies, bunnies, calves)	Boys	10	16	4	30
	Girls	12	8	3	23
4. Where do we get babies?	Boys	14	8		22
	Girls	7	4		11
5. Where did Mrs. M—— get the baby?	Boys	8	6		14
	Girls	20	2	1	23
6. Where did I come from?	Boys	18	8		26
	Girls	17	6		23
7. Where was I before I was here?	Boys	2			2
	Girls	5			5
8. Where was I before I was born?	Boys	1			1
	Girls	4	1		5
9. Where was I when you were a little girl?	Boys	3			3
	Girls	2			2
10. Was I at your wedding? Where was I then?	Boys	6	3		9
	Girls	8			8
11. Did you really find me in a corn field?	Boys	2	2		4
	Girls	4	2		6
12. How much did I cost?	Boys	3	3	1	7
	Girls	1	1		2
13. Do babies really come from storks? Santa Claus? Baby-land?	Boys	4	4		8
	Girls	1	5		6
14. How can a stork carry a baby when the baby is so heavy?	Boys	1	3	1	5
	Girls		3	1	4

TABLE 7 (*continued*)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
15. Why doesn't the stork bring the baby to the door instead of to the hospital?	Boys Girls	2	1		3
16. Mother, why did you say the stork brought me?	Boys Girls		2		2
17. Why do you always go to the hospital for these babies?	Boys Girls	1 5	2 4		3 9
18. Where does the hospital get the babies?	Boys Girls	4 4	2 2	1	7 6
19. How did you happen to pick me out?	Boys Girls	1 1	2	1	4 1
20. How do the doctors get the babies at the hospital?	Boys Girls	7 9	2 1	1	10 10
21. Does God give us all the babies?	Boys Girls	3 1	1		4 1
22. If God makes babies, who made God?	Boys Girls	2	3		5
23. Who took care of the first baby?	Boys Girls	1	4		5
24. How are babies made?	Boys Girls	5 5	3 1	1	9 6
25. How does a baby get down from Heaven if God makes them?	Boys Girls	8 3	3		11 3
26. How do we begin?	Boys Girls	1			1
27. How do the babies come?	Boys Girls	4 1	6 1		10 2
28. How do babies grow?	Boys Girls	1 4	1 2		2 6
29. How do horses get their babies? Birds? Puppies?	Boys Girls	3 1	8	1	12 1
30. Did I grow from a little seed?	Boys Girls	3	2 1		5 1
31. Mother, did we come from an egg?	Boys Girls	1 1	1 1		2 2
32. Do you have to sit on the egg?	Boys Girls	1			1
33. Do cows and horses lay eggs?	Boys Girls	1	3	1	5
34. Why don't you tell me what I ask? Why do you always tell me lies?	Boys Girls		2		2
35. Won't you tell brother what you have told me about babies?	Boys Girls		1		1

TABLE 7 (continued)

Questions		Ages			No. of
		2 to 5	6 to 9	10 to 13	questions
36. What is it about babies? (Discussion overheard among playmates)	Boys				
	Girls		1		1
All	Boys	215	157	19	391
	Girls	207	107	17	331
		422	264	36	722
Cumulative per cent of 722 questions		58.45	36.56	4.99	100.00
<i>Group 2: 256 questions concerning the coming of another baby</i>					
1. Mother, why can't we have a baby?	Boys	11	7	3	21
	Girls	12	14	4	30
2. How can we get a baby?	Boys	1			1
	Girls	1			1
3. Where can I pick out the baby?	Boys	1			1
	Girls	1			1
4. Have Tom and I enough money to buy a baby?	Boys	6	6		12
	Girls				
5. Why do we buy so many?	Boys		1		1
	Girls				
6. Did the stork come into our yard today?	Boys				
	Girls		2		2
7. If I should pray every night would I get a baby sister?	Boys	1			1
	Girls			1	1
8. Is the doctor going to bring us the baby?	Boys	2	2		4
	Girls				
9. Why don't you get a baby from the hospital?	Boys	1	2	2	5
	Girls	2	7		9
10. Why don't you get sick and go to the hospital and have a baby?	Boys				
	Girls		4		4
11. But wasn't there a baby when you were so sick?	Boys				
	Girls	2	1		3
12. Where are you going to get the new baby?	Boys	6	3		9
	Girls	4	2		6
13. What makes Mrs. S—— so fat?	Boys	13	14	5	32
	Girls	4	6	6	16
14. Why isn't she a nice shape?	Boys				
	Girls	6	2	1	9
15. Is it true that women get big before they have babies?	Boys		1	1	2
	Girls		1	3	4
16. How do you know that the cat is going to have kittens?	Boys	3	10	1	14
	Girls	1	4		5
17. Is Auntie going to have a baby?	Boys		1	3	4
	Girls				

TABLE 7 (*continued*)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
18. What are you making those clothes for?	Boys	2	1	1	4
	Girls	3	1		4
19. What is a baby shower?	Boys		2		2
	Girls				
20. Is it true that we are going to have a baby?	Boys		1		1
	Girls	1	2		3
21. Where is our little sister?	Boys	1			1
	Girls	3			3
22. How do you know that we are going to have a baby?	Boys			2	2
	Girls	3	2	1	6
23. When is the baby coming?	Boys				
	Girls		3		3
24. Isn't it funny they know it is coming so much ahead?	Boys			1	1
	Girls				
25. Why doesn't the baby hurry up and come?	Boys	2	2		4
	Girls		1		1
26. What if it doesn't match? (the family)	Boys	1			1
	Girls				
27. Who does the baby belong to?	Boys		2		2
	Girls				
28. Why do you say "her," Mother?	Boys		1		1
	Girls				
29. Why couldn't it have been a boy?	Boys	2	4		6
	Girls	1	2		3
30. Why did you get a tiny baby?	Boys		1		1
	Girls				
31. Why do people have new babies?	Boys	1			1
	Girls				
32. What do they do with the old ones?	Boys	1			1
	Girls				
33. If a Negro man and his wife went to China would they have to get a Chinese baby?	Boys			1	1
	Girls				
34. Why did Mrs. Jones pick out that kind of a baby? (crippled)	Boys		1		1
	Girls				
35. Why didn't the stork bring a horse instead of a little brother?	Boys	1			1
	Girls			1	1
36. How does it happen that — have four babies and — only one?	Boys		1		1
	Girls				
37. Grandmother's cat is going to have five kittens, isn't she? She has five bumps.	Boys				
	Girls	1			1

TABLE 7 (continued)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
38. (Seeing puppies just born)	Boys	1			1
Why, Trixie has had a birthday, hasn't she?	Girls				
All	Boys	57	63	20	140
	Girls	45	54	17	116
		102	117	37	256
Cumulative per cent of 256 questions		39.84	45.70	14.46	100.00
<i>Group 3: 42 questions concerning intra-uterine growth</i>					
1. Do babies really grow inside the mother?	Boys		4		4
	Girls		4	1	5
2. How does a baby grow in its mother's stomach?	Boys	1	3	1	5
	Girls	5	3	1	9
3. How did the egg get inside the chicken?	Boys				
	Girls	1		1	2
4. Did the cat carry the kittens like a kangaroo does?	Boys		1		1
	Girls	1			1
5. How do you know that the baby is inside of you?	Boys				
	Girls	1		1	2
6. Do you 'spose I have a baby in me?	Boys	1			1
	Girls		1		1
7. How does a baby breathe before it is born? How does it eat?	Boys	1	1		2
	Girls				
8. Are our arms or legs put on our body first?	Boys	1			1
	Girls				
9. How long will it take the baby to grow?	Boys		3		3
	Girls		2		2
10. Won't the baby kick her and hurt her?	Boys				
	Girls		1		1
11. When you open the egg how do you know that there won't be a baby in it?	Boys				
	Girls	1			1
12. Do you think I could tell Nancy about the cat and how her kittens grow?	Boys		1		1
	Girls				
All	Boys	4	13	1	18
	Girls	9	11	4	24
		13	24	5	42
Cumulative per cent of 42 questions		30.95	57.14	11.91	100.00
<i>Group 4: 183 questions concerning the process of birth</i>					
1. What is "born"?	Boys	1			1
	Girls				

TABLE 7 (*continued*)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
2. How do you know when the baby is going to be born?	Boys		1	1	2
	Girls				
3. Why do mothers go to the hospital?	Boys	6	5		11
	Girls	3	6	1	10
4. Why does the mother stay in the hospital?	Boys	3	5	1	9
	Girls	3	3	4	10
5. Why are mothers sick when they go to the hospital to get a baby?	Boys	5	6		11
	Girls	4	2		6
6. Why are there doctors when babies are born?	Boys				
	Girls		2		2
7. Did the baby hurt you when it was born?	Boys	1	2		3
	Girls		2		2
8. How does the baby get out of the mother?	Boys	6	16	5	27
	Girls		11	2	13
9. Where did the baby come out?	Boys		3		3
	Girls		2	1	3
10. After they are grown do they come out of your mouth?	Boys	1		1	2
	Girls		1		1
11. Don't you have to burst open when you have a baby?	Boys		1		1
	Girls		1		1
12. Does the doctor cut a round hole in the mother's stomach?	Boys				
	Girls	1	1		2
13. Your stomach is smaller now, isn't it?	Boys	1			1
	Girls	5			5
14. Is that the way they come out, head first?	Boys		1	1	2
	Girls			1	1
15. Are kittens, dogs, and babies laid like chickens?	Boys	3	5	1	9
	Girls	2	3	1	6
16. Are babies born like puppies?	Boys	1	4		5
	Girls	2	2		4
17. Do dogs hatch puppies?	Boys			1	1
	Girls	1	1		2
18. How do cows get born?	Boys		1		1
	Girls				
19. The next time we have a baby can I see how it is born?	Boys		1		1
	Girls				
20. Are babies always black when they are born?	Boys		1		1
	Girls			2	2
21. About how big was I when I was born?	Boys	1	3		4
	Girls		1		1
22. Who saw me first?	Boys		1		1
	Girls				
23. Did you like me awfully well, Mother, when I was born?	Boys		1		1
	Girls				

TABLE 7 (continued)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
24. What does it mean—some are born and some come from eggs?	Boys Girls	1	1		1 1
25. Where does the egg come from?	Boys Girls	1 1		2	3 1
26. Which comes first, the chicken or the egg?	Boys Girls	1			1
27. Which came first, Mary or me? (sister younger)	Boys Girls				
28. Why didn't I hatch like a bird's egg?	Boys Girls	1 1		1	2 1
29. Why are there big birds and the little birds at one time?	Boys Girls				
30. What did you tell me that for about the seed? It wasn't so.	Boys Girls				
All	Boys Girls	32 29	57 39	14 12	103 80
		61	96	26	183
Cumulative per cent of 183 questions		33.33	52.46	14.21	100.00
<i>Group 5: 209 questions concerning organs and functions of the body</i>					
1. What is a cell?	Boys Girls			1	1
2. What is a navel?	Boys Girls	3 5	2	1	6 5
3. What are those?	Boys Girls				
4. What are they called? (breasts)	Boys Girls	3 2	2		5 2
5. Why are your breasts so much bigger than mine?	Boys Girls	4	7	1	12
6. What are your breasts for?	Boys Girls	1 4	1	1	3 4
7. Is the baby biting you?	Boys Girls	8	3	1	12
8. And is that other one the baby's too? (mother's breast)	Boys Girls				
9. Is that where you used to feed me?	Boys Girls		1		1
10. Why do you have milk?	Boys Girls		2 1		2 2
11. Why haven't I milk like you, Mother?	Boys Girls	1			1

TABLE 7 (*continued*)

Questions	Ages			No. of questions
	2 to 5	6 to 9	10 to 13	
12. Where does the milk come from?	Boys Girls	3		3
13. How can the baby get carrots and beets through your breast?	Boys Girls	1		1
14. Why can't you feed S——'s baby?	Boys Girls	1		1
15. Why couldn't we give her milk that we buy?	Boys Girls		1	1
16. Won't I ever have milk so that I can feed the baby?	Boys Girls	1		1
17. Does a colt live on its mother's milk?	Boys Girls	1	1	2
18. Why do you have hair on your body?	Boys Girls	4 9		4 15
19. What is this? (penis)	Boys Girls	8 14	2	10 18
20. Why do I urinate?	Boys Girls	2 1		2 1
21. Mother, have pigs lots of places to wet?	Boys Girls	1		1
22. Why can't we urinate out under the lovely trees like on picnic days?	Boys Girls	1		1
23. Why do all the dogs and horses and things do it outside then?	Boys Girls	1		1
24. Why can't I go to the toilet with Harry?	Boys Girls	1 3	1	5
25. Why does J——'s penis stand up when he doesn't have to urinate?	Boys Girls	1	1	2
26. Why do I have to keep myself so clean?	Boys Girls	3 1	1	4 1
27. Why am I different from other boys? (circumcised)	Boys Girls		1	1
28. Will A—— hurt himself if he handles himself like that?	Boys Girls		1 1	2
29. Mother, what was the boy doing that they punished him? (masturbating)	Boys Girls		1	1
30. Why do those boys think it funny to talk about the things all of them have?	Boys Girls		1 1	1 2
31. What is this other thing for? (testicles)	Boys Girls	3 2		5

TABLE 7 (continued)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
32. Does Papa have them, too? (penis and testicles)	Boys Girls	1			1
33. When I get big can I be a daddy?	Boys Girls	1			1
34. Can I have ten children when I get to be a mamma?	Boys Girls	1 1	3 1		4 2
35. Why doesn't everybody have a baby?	Boys Girls				2
36. What is the womb?	Boys Girls	1 1		1	2 2
37. What is Kotex?	Boys Girls		2 1	3 6	5 9
38. Why do some girls wear belts?	Boys Girls		2	1	5
39. Mother, did you hurt yourself? (spot of blood upon clothing)	Boys Girls			1	1 3
40. Do girls bleed when they get older?	Boys Girls		3		4
41. What in the world is happening to me? (menstruation)	Boys Girls			1	1
42. Can't you tell me tonight what it is that happens to girls at 13?	Boys Girls	1	1		2
43. Will I have that, too? (menstruation)	Boys Girls	1	1		2
44. How long does that menstruation last?	Boys Girls		1		1
45. What is menstruation for?	Boys Girls			3	3
46. What does "female" mean?	Boys Girls		1		1
47. What does "pregnant mother" mean?	Boys Girls			1	1
48. When sister gets grown up we will have two mammas, won't we?	Boys Girls		1		1
49. What does the dog have to be operated for?	Boys Girls		1		1
50. Do they really operate on dogs so they can't have any puppies?	Boys Girls		1 1		1 2
All	Boys Girls	46 61	40 30	18 14	104 105
		107	70	32	209
Cumulative per cent of 209 questions		51.20	33.49	15.31	100.00

TABLE 7 (continued)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
<i>Group 6: 226 questions concerning physical sex differences</i>					
1. How can you tell when the baby comes whether it is a boy or girl?	Boys	1	5	1	7
	Girls		1	1	2
2. Why is Alice a girl and John a boy?	Boys	2	1		3
	Girls	4	3		7
3. How can you tell the difference between boys and girls?	Boys		3		3
	Girls	1	5	1	7
4. What is the difference between a cow and a bull?	Boys		1		1
	Girls	1	4	4	9
5. What is this growing on this little kitten?	Boys				
	Girls		2		2
6. Mother, are you sure it isn't a boy?	Boys				
	Girls		1		1
7. Why don't you put a dress on brother and then there would be three girls in our family?	Boys				
	Girls		1		1
8. Are boys and girls built alike?	Boys	2	2		4
	Girls	1	3		4
9. Why can't I come in with you, Mother?	Boys	6	2	1	9
	Girls	1			1
10. You are different than I am, aren't you, Mother?	Boys	6			6
	Girls				
11. Is sister (or brother) made like you and brother (or sister) made like me and Daddy (or Mother)?	Boys	3			3
	Girls	2			2
12. Why are fathers and mothers different?	Boys	1	1		2
	Girls	3			3
13. Why are you different from me, Daddy?	Boys				
	Girls	6	2		8
14. Why does Daddy stand up and you sit down? (urination)	Boys	3			3
	Girls	5	1		6
15. Is the baby hurt? (sister—external sex organs)	Boys	3			3
	Girls				
16. She (or he) isn't like me, is she (or he)?	Boys	10	4		14
	Girls	1			1
17. Why isn't sister (or brother) like me?	Boys	11	6		17
	Girls	18	7		26
18. Sister is going to grow like me after while, isn't she?	Boys	3			3
	Girls				
19. Why do little girls sit down on the toilet?	Boys	11			11
	Girls				
20. What's that? (brother's penis)	Boys				
	Girls	5	3		8

TABLE 7 (continued)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
21. He's a boy, isn't he?	Boys				
	Girls	9			9
22. Why doesn't she go to the toilet the way I do?	Boys	6	2	2	10
	Girls				
23. Why do men have hair on their bodies?	Boys	3	1		4
	Girls				
24. Will I be all hair when I grow up?	Boys		1	1	2
	Girls	3		2	5
25. Why doesn't Daddy have big breasts like you do, Mother?	Boys	3	1		4
	Girls	2	1		3
26. Why can't fathers give milk to the babies?	Boys				
	Girls	1	1		2
27. Why can't our dog have puppies? (male)	Boys		3	1	4
	Girls		2	1	3
All	Boys	74	33	6	113
	Girls	63	37	10	110
		137	70	16	223
Cumulative per cent of 223 questions		61.43	31.39	7.18	100.00
<i>Group 7: 92 questions concerning relation of father to reproduction</i>					
1. Why do you have to have a daddy?	Boys		1		1
	Girls				
2. Do boys come from the daddy and girls from the mother?	Boys				
	Girls	1			1
3. What do you do to get a baby?	Boys				
	Girls		1		1
4. How does the baby get started?	Boys		1	1	2
	Girls				
5. How did the baby get in there?	Boys	2	1	1	4
	Girls	1	1	1	3
6. If the baby grows in the mother why does it look like the father?	Boys		2		2
	Girls				
7. What does the father have to do with it?	Boys			1	1
	Girls				
8. Who plants the seeds for the baby?	Boys	1			1
	Girls				
9. How does the father part get into the mother?	Boys		2	1	3
	Girls				
10. How do the seeds get into the mother cat?	Boys				
	Girls		1		1
11. Just how does the father plant the seed in the mother?	Boys			2	2
	Girls		1	2	3

TABLE 7 (*continued*)

Questions		Ages			No. of questions
		2 to 5	6 to 9	10 to 13	
12. How does the father know when to plant the seed?	Boys Girls		1		1
13. Why did they do that if they could not afford another baby?	Boys Girls			1	1
14. Can anyone have a baby after the father is dead?	Boys Girls		1		1
15. How can the baby be my brother when Carl (step-father) wasn't my father?	Boys Girls	1		1	2
16. The cat had kittens and there wasn't any father, was there?	Boys Girls		1 2		1 3
17. How do you know that the little cats came from the little boy cats?	Boys Girls		1		1
18. Why are you putting the birds in the same cage?	Boys Girls		3	1	4
19. What are those dogs doing?	Boys Girls	2 1	4 2	2 1	8 4
20. Mother, those dogs are getting married, aren't they?	Boys Girls			1	1
21. Why do roosters run after hens?	Boys Girls		3		3
22. What do you mean by mating?	Boys Girls	2 1	3	1	6 1
23. What is the bull for?	Boys Girls	2 2	2	3 3	7 5
24. What does the father rabbit put into the mother to make the baby rabbit?	Boys Girls		1		1
25. What does that term mean, Mother? (street term for intercourse)	Boys Girls	1	3 4	3 1	7 5
26. Can't I ever have a baby?	Boys Girls	1	1		2
27. How can you know when a boy is grown up to be a daddy?	Boys Girls		1		1
All	Boys Girls	11 8	24 21	17 11	52 40
		19	45	28	92
Cumulative per cent of 92 questions		20.65	48.91	40.44	100.00

TABLE 7 (continued)

Questions		2 to 5	Ages 6 to 9	10 to 13	No. of questions
<i>Group 8: 36 questions concerning marriage</i>					
1. What is "married"?	Boys	1			1
	Girls				
2. All three of us are married, aren't we?	Boys				
	Girls	1	1		2
3. Why do people marry?	Boys				
	Girls	1	1		2
4. Did you have to marry Daddy?	Boys				
	Girls		1		1
5. Was I born before you were married or after?	Boys				
	Girls		1		1
6. How does God know when people are married?	Boys				
	Girls		1		1
7. Do you have to be married to have a baby?	Boys		1	1	2
	Girls		5	1	6
8. Why do only married people have children?	Boys	1	1		2
	Girls		1		1
9. Do you have to be married every time you have a baby?	Boys		1		1
	Girls		1		1
10. Why do women have to be married before they have a baby?	Boys			1	1
	Girls			1	1
11. When I marry will I have babies?	Boys			1	1
	Girls				
12. How can she have a baby if she isn't married?	Boys				
	Girls		1		1
13. How can she have puppies if she isn't married?	Boys				
	Girls		1		1
14. Why can't unmarried girls have babies?	Boys				
	Girls		1	1	2
15. Why would it be wrong for me to have a baby?	Boys				
	Girls			1	1
16. Why should that girl be turned outside just because she had a baby?	Boys			1	1
	Girls				
17. Why do unmarried couples have to have chaperones when staying at the same place?	Boys				
	Girls			2	2
18. Can cousins marry?	Boys		1		1
	Girls				
19. Do you think the cat wants to get married to our cat?	Boys			1	1
	Girls		1		1
All	Boys	2	4	3	9
	Girls	2	17	8	27
		4	21	11	36

TABLE 7 (*continued*)

Questions		2 to 5	Ages 6 to 9	10 to 13	No. of questions
Cumulative per cent of 36 questions		11.11	58.33	30.56	100.00
All questions	Boys	441	384	98	923
	Girls	424	323	93	840
		865	707	191	1763
Cumulative per cent of 1,763 questions		49.06	40.10	10.84	100.00

RESPONSIBILITY BEING ASSUMED BY PARENTS FOR GIVING SEX INSTRUCTION

It is of interest to review here the background and reactions of the 981 mothers who helped with this study through contributing questions and information concerning themselves and their children.

The largest group, about 40%, were between the ages of thirty and thirty-five years as estimated at the time of the interview. About 20% were from twenty-five to thirty years old, and about 20% were between the ages of thirty-five and forty.

The educational background, as given by the mothers, showed that 17.8% were college graduates and that 45.5% had attended high school.

The mothers' own sources of early and later information upon matters of sex were miscellaneous in character. There were 152, or 15.5%, who received early information at home, and 343, or 35.0%, who mentioned "playmates" as their early source of knowledge. Frequent combinations which were given were "playmates, no later information," "playmates and observation," "playmates and

TABLE 8
EDUCATIONAL BACKGROUND OF 981 MOTHERS

	Number	Percentage
Grade school	205	20.9
High school	447	45.5
Business college	31	3.2
Normal school	70	7.1
College	171	17.5
Postgraduate	3	.3
Not stated	54	6.0
Total	981	100.0

TABLE 9
SEX INSTRUCTION INITIATED BY MOTHERS

Type of instruction	Number of mothers attempting instruction
1. Reproduction	
Mother's part	388
Father's part	92
Mother's part in animals	97
Father's part in animals	53
2. Sex differences	547
3. Puberty	
Menstruation	107
Seminal emissions	11
4. Vocabulary—organs and functions	67

husbands," and "parent and experience." Fifty-four, or 5.5%, had received later information from adults, 105, or 10.7%, had help from books or pamphlets, while 110, or 11.2%, claimed to have had nothing beyond their own experience in childhood to aid them in training their children in reference to matters of sex.

Notwithstanding their own lack of background and methods for procedure, as stated by themselves, these mothers were conscious of their children's curiosity and welcomed, in most instances, the visit of the parent adviser. They reported at least the beginning of an effort on the part of parents to meet these needs of children.

There were 167 mothers who had obviously given evasive or untruthful information. Nine had given some instruction, though they were not specific as to its nature. There were 204 mothers who frankly admitted that they had given no instruction. There has since been made available to these mothers a local conference course in their respective neighborhoods so that every mother may, if she desires, avail herself of the newer methods and materials which are now being offered in this field.

SUMMARY AND CONCLUSIONS

This study was undertaken for the purpose of studying children's questions in regard to sex in the early years as shown by questions gathered from parents whose children were under junior-high-school age.

Questions of children between the ages of two and fourteen years were accumulated as one item of information in the house-to-house interviews held with mothers by the parent advisers in the Women's

Co-operative Alliance, Minneapolis, Minnesota, from April 16, 1926, to January 1, 1928, these individual contacts being a definite part of the sex-education program of that organization. Children in each family were noted as to ages and sex. The questions, with the ages of the children at the time the questions were asked, were recorded as reported by parents. Nine hundred eighty-one homes with 1797 children are included in the study. There were 1763 questions reported.

There were 563 boys and 546 girls for whom questions were reported. These numbers represent 61.9% of the boys and 61.5% of the girls. There was, apparently, an almost equal interest displayed in the subject through the medium of questions by boys and girls.

In comparing the boys and girls who had passed through the age at the time of the interview, it was found that the number of boys who asked questions was quite evenly distributed from four to twelve years, with slight peaks occurring at five and ten years; the number of girls who asked questions was distributed, though less evenly, from four to thirteen years, with peaks occurring at five, ten, and thirteen years. The percentage of the boys who asked questions varied from 14.5 at four years to 21.8 at ten years; the percentage of the girls who asked questions varied from 10.0 at twelve years to 21.6 at thirteen years.

When the 1763 questions were classified according to the ages at which they were asked, the greatest number of questions for boys occurred at five (1.44 questions), and at nine, (1.57 questions); for girls, at five (1.41 questions), at nine, (1.41 questions), and at thirteen years (1.50 questions).

The first questions of boys came chiefly between four and eleven years, with 16.1% at five years and 16.7% at ten years. The first questions of girls came between four and thirteen years of age, with 17.9% at four years, 15.6% at ten years, and 17.8% at thirteen years.

The 1763 questions may be classified into eight groups; origin of babies, 40.9%; coming of another baby, 14.5%; physical sex differences, 12.7%; organs and functions of the body, 11.9%; process of birth, 10.4%; relation of the father to reproduction, 5.2%; intra-uterine growth, 2.4%; and marriage, 2.0%.

The questions are tabulated by groups, according to age levels, in years, for boys and for girls. For analysis of the interests of children as displayed in the subject-matter of the questions, the questions are considered for preschool children, two to five years, who asked

49.1% of the questions; for children of early school age, six to ten years, who asked 40.1% of the questions; and for pre-adolescent children, ten to thirteen years, who asked 10.8% of the questions. It must be borne in mind, however, when considering these percentages that all of the children reported to have asked questions, that is, 100%, were either in the preschool age group or had passed through it; while only 22% of the children had arrived at the pre-adolescent age. There were, therefore, more questions reported for the six to nine years age level than for either the preschool or the pre-adolescent periods.

When questions of the children of preschool age are studied alone, the sex interests, ranked, are found to center in questions represented by the groups: origin of babies, physical sex differences, organs and functions of the body, and the coming of another baby.

When questions of children of early school age are studied alone, the sex interests of these children center in the origin of babies, the coming of new babies, organs and functions of the body, with added interest in the process of birth. The questions of the children of pre-adolescent age show a decline in the interests of the preceding years. Knowledge sought in these years is represented by the group of questions concerning the father's relations to reproduction and marriage.

The relatively large number of questions for children of preschool age, as compared with the other ages, indicates that children are manifesting interest in sex through their questions in these very early years. The subject-matter of the questions indicates what children are wanting to know in reference to sex. Interest is shown to develop from simple inquiries concerning the origin of babies, through the more intricate physiological processes of conception and birth, to the deeper sociological and psychological interpretations of customs and attitudes.

The ages at which the questions occur substantiate the feeling expressed by so many writers that sex instruction should be given in the early years. The questions, in showing the trend of children's interest in sex, seem to indicate a developing interest in the subject which points to the need of a graded program in sex education for children.

*The State University of Iowa
Iowa City, Iowa*

UNE ÉTUDE DES QUESTIONS DES JEUNES ENFANTS À PROPOS
DE SEXE: UNE PHASE D'UNE APPROCHE EXPÉRIMENTALE À
L'ÉDUCATION DES PARENTS

(Résumé)

Cette étude essaie de répondre à trois questions: (1) Les jeunes enfants en général posent-ils des questions à propos de sexe? (2) S'ils posent celles-ci, à quels âges spécifiques viendront les questions? (3) Quels renseignements les enfants cherchent-ils?

Pendant une période de dix-sept mois on a accumulé 1763 questions des enfants âgés de 2 à 14 ans, ces questions étant un de plusieurs renseignements obtenus par les conseillers des parents de la Women's Cooperative Alliance, Minneapolis, Minnesota, dans des entrevues de maison en maison avec 981 mères en 14 circonscriptions scolaires.

On a rapporté des questions pour 563 garçons et 546 filles. Ces nombres représentent 61,9 pour cent des garçons et 61,5 pour cent des filles. Les garçons ont posé le plus grand nombre de questions à cinq ans et à neuf ans, et les filles à cinq ans, à neuf ans, et à treize ans. Les premières questions rapportées pour les garçons ont été pour la plupart entre les âges de quatre ans et de onze ans avec de petites hauteurs à cinq ans et à dix ans; et pour les filles entre les âges de quatre ans et de treize ans avec de petites hauteurs à quatre ans, à dix ans et à treize ans.

Les 1763 questions tombent en huit groupes: origine des bébés, 40,7 pour cent; arrivée d'un nouveau bébé, 14,5 pour cent; différences physiques de sexe, 12,7; organes et fonctions du corps, 11,9; procédé de la naissance, 10,4; relation du père à la reproduction, 5,2; croissance intra-utérine, 2,4; et mariage, 2,0 pour cent.

HATTENDORF

EINE UNTERSUCHUNG DER FRAGEN JUNGER KINDER ÜBER DAS SEXUELLE: EINE PHASE EINER EXPERIMENTELLEN ANNÄHERUNG AN DIE ELTERLICHE ERZIEHUNG

(Referat)

In diesser Untersuchung versucht man, drei Fragen zu beantworten: (1) Stellen kleine Kinder im Allgemeinen Fragen über das Sexuelle (about sex)? (2) Wenn ja, in welchen besonderen Altern werden die Fragen wahrscheinlich kommen? (3) Ueber welche Tatsachen suchen Kinder sich zu erkundigen?

Es wurden 1763 Fragen von Kindern im Alter von 2 bis 14 Jahren über 17 Monaten gesammelt. Diese Fragensammlung war ein Bestandteil der in einer Serie von Gesprächen (interviews) mit 981 Müttern von Haus zu Haus gesammelten Daten, die, in 14 Schulbezirken, von Elternberatern (parent advisers) der Women's Cooperative Alliance, Minneapolis, Minnesota, zusammengebracht wurden.

Es wurden Fragen von 563 Knaben und 546 Mädchen notiert. Diese Zahlen stellen 61.9% der Knaben und 61.5% der Mädchen dar. Die grösste Aufhäufung der Fragen fand bei Knaben mit 5 und mit 9 Jahren, und bei Mädchen mit 5, 9, und 13 Jahren statt. Die ersten für die Knaben notierten Fragen kamen meistens zwischen 4 und 11 Jahren, und häuften sich bei 5 und bei 10 Jahren etwas auf. Bei Mädchen kamen sie besonders zwischen 4 und 13 Jahren und häuften sich bei 4, 10, und 13 Jahren etwas auf.

Die 1763 Fragen lassen sich in 8 Gruppen teilen: der Ursprung von kleinen Kindern (babies), 40.9%; die Ankunft eines neuen Kindes [in der Familie] 14.5%; physiologische sexuelle Unterschiede, 12.7%; Organe und Funktionen des Körpers, 11.9; der Geburtsvorgang, 10.4; die Beziehung des Vaters zu der reproduction, 5.2; das Wachstum innerhalb der Gebärmutter, 2.4, und die Vermählung, 2.0%.

HATTENDORF

THE PROBLEM OF CRIMINALITY IN CHILDREN

From the University of Bologna

G. C. FERRARI

It is very interesting to observe the differences between the American and the European practices in the study and practice in the general field of criminology.

In Italy, as early as 1761, there appeared the well-known study of Cesare Becarria, *Dei Delitti e delle Pene*, in which the author clearly demonstrated the uselessness of, and the harm caused by, punishment and the possibility of moral prophylaxis for crime. In 1876 Cesare Lombroso published his famous *Uomo Delinquente*, a book which should have demonstrated the biological necessity of crime. But these volumes, though they deeply interested the cultural world, had no practical influence in Italy, although they have had a marked influence on the penal legislation of most other countries.

The lack of a scientific attitude is especially evident in the case of youthful criminals. Whereas Lombroso emphasized that responsibility for crime should be weighed according to the physical and mental personality of the criminal, in Italy the imputability of children begins at the age of nine years, while practice and applied psychology have led English jurists (Children's Act, 1908) to believe that it is only at the age of fourteen that children can be assumed to have reached the *quantum* of discernment which may give rise to their imputability. And legislative practices of other modern countries tends toward the heightening of this limit, setting it always nearer the coming of age. In America, especially, we find this tendency very strongly at work. In fact, America has preceded by some score of years all other nations in providing for her youthful delinquents by way of looking after their moral and physical education and creating for them special institutions and organizations which have become world famous.

She has, however, done this by impulse. It is only relatively recently that America has given any scientific contribution to the problem of child criminality. The names of W. I. Thomas, C. N. Sutherland, Jane Addams, Miriam van Waters, and a few others

stand out in this recent work. It almost seems as if these eminent scholars had sought to find a scientific justification for all that America has already done, leading, as she does, all the nations of the world in this work.

The present paper deals with the general problem of child criminality or delinquency and the means by which it can best be considered. The problem is a serious one for two reasons: first, because youthful criminality, for the greater part, automatically generates adult criminality; and, secondly, because the current ideas on youthful criminality are far from clear.

If we observe directly, without prejudice or partiality, it appears that the problem of "youthful criminality" is more or less of the type which the Germans call a "*Scheinproblem*": That which we call child criminality, as a psychological fact, arises especially from our unfortunate lack of comprehension of the psychology of the young, who are too far away from us, either in years or in social conditions, and who are observed during a period of evolution, a period which is therefore turbid and confused. As a social fact, the problem arises from presumed judicial necessities.

The psychological problem is well revealed in two methods of dealing with offenders. We are all ready to deplore the far from brilliant effects obtained by the reformatories that pretend to transform the mind and the heart of youth, but at the same time we are all equally ready to forget the remarkable results obtained by the treatment of misguided children when undertaken by such men as Don Bosco, Dr. Barnardo, and so many others who devote themselves to this problem. These men collect around them, without any choice or any reservation, *all* the youths they meet who have gone astray or been forsaken; yet they never speak of criminals, nor do any criminals come out of their schools or workshops.

The difference between the results obtained by the reformatories and those brought about by these men lies in the fact that the reformatories are bureaucratic offices and therefore inevitably rigid, while the children are living beings who feel and suffer each one in his own way. Adaptation and understanding are practically impossible.

Thirty years' experience with children has taught me that the natural activity of children when left to themselves is simply the direct expression of their "instinctive tendencies." The commonest

among these are (according to Warren) imitation, play, curiosity, a rudimentary tendency towards esthetic expression, association with others, etc. In some cases we also see the outcropping of certain of the instincts, such as that of nutrition, of reproduction, of defense, etc., modified by their surroundings and more or less inhibited by intellectual circumstances.

If we observe the poor children who are not looked after and who go about in groups in the suburbs of our towns, we find them pilfering, hiding in caves or abandoned huts, displaying their ability in breaking windows by throwing stones, ill-treating any animal they may be able to catch, choosing a chief whom they obey, etc. And when their continued operation in a given place has led to the intervention of the law, they shift their headquarters, and go right on doing the same things. That such activities, which are necessarily anti-social at best, should lead to acts of violence is but natural.

Children of the upper social classes have the same tendencies as have these more unfortunate children. If we find that they present no such problems to society, it is not because of "preaching" or of the "good example" of which we read in educational literature. It is only because all the practice and the habits of these well-to-do children tend towards the satisfaction of their instinctive tendencies as they arise so that they can be bridled, masked, or symbiotized, and thus prevented from becoming anti-social.

Thus, though our children do not have the stimulus of hunger, we prevent them from stealing the fruit or sweets of others by allowing them to buy their own; we take them out for walks, allow them to participate in sports to satisfy their desire to struggle and to distinguish themselves; by means of moving pictures, museums, theatres, exhibitions, etc., we satisfy their curiosity, their spirit of adventure; we help them to make collections, thus satisfying their instinct of acquisition and of property; by hunting or shooting or even by experimental vivisection we satisfy the desire which is in all of us to be cruel, to do harm.

Such familial education in the well-to-do classes tends to confirm the idea that tendencies which may become anti-social are natural in children. It is merely a question of money if, with the rich, the opportunity of satisfying, at the right time and in the right way, these instinctive tendencies prevents them from developing or from

having harmful effects or creating anti-social habits. With poor children such opportunities are utterly lacking.

Another indication of the naturalness of their "crimes" is found by questioning the children of these poorer classes when they have come to trial for such crimes. Their answers, when circumstances do not alter them but allow them to be honest, always tend to display the naturalness of the actions and thoughts.

From my own personal experience, I should say that the child never has the conception of the general value of the immoral act that he commits. He knows that the act itself is considered wrong, but he justifies it, putting it into direct relation, from effect to cause, with a suitable stimulus. If, for instance, his violence has not been merely a manifestation of his exuberant vitality, he will always be able to say to which stimulus—with either real or exaggerated appreciation—he intended to respond. It is only later, when the constant and successful exercising of his strength has made him overbearing that he will naturally like to make use of the argument *quia nominar leo*—at which no one can be surprised.

Theft occurs in the same way. At first the child "takes" other people's things; but one cannot say that he is stealing because the act is committed without any scruple and without that slyness that would be necessary to escape the first consequences of the act itself; and it is clear that to the child it is a contingency of *no importance whatever* whether the thing he takes is his or belongs to someone else. But if his attention is called to the fact that he has taken the property of another he will *try to find a "reason"*—but not an *excuse*—merely to satisfy the strange curiosity of whoever may be interested in such a question. But it is always a "reason" due to circumstances—to a true *fin de non recevoir*—expressed to please the questioner: the things were left out, not put away; the shoes fitted him so well; there were so many. . . ., but this not to excuse himself, since this would be utterly superfluous and out of place.

It is the constant recurrence of this mechanism and its generality that indicates how natural and genuine are these answers—also the fact that they are always restricted to the case in question, the child never thinking of extending any explanation to similar cases. Abstraction does not occur outside the empty verbalism of his school book or his religious texts.

No doubt, some outside conditions favor this attitude, which we

may call individualistic. In the first place, only a few of the children's crimes are found out and punished, so that for the others their evaluation is only subjective. Furthermore, the variety of the verdicts—which differ according to the seriousness of the crimes, or their repetition, etc.—gives the child the idea that “it is all a question of luck.” And thus it is always more difficult for the child to reach, in a general way, the abstract conception of punishment, and still more difficult for him to reach the height of the general idea of his fault. This is so true that children feel the necessity of individualizing Justice—with such and such a judge “who has a tooth against us boys,” or with a policeman “who always puts the blame on me”—without ever being capable of imagining which are the true social functions of judges or of the police.

If a child breaks a street-lamp, it is certainly not with the idea of doing any special harm to the proprietor or to the town council. He is merely practicing his aim, and the satisfaction of a well-directed aim is so great that, even were he to succeed in understanding the harm done to others, it would be difficult for this idea to overcome the pleasure given him by his well-aimed stone. Now, if the child cannot grasp this antagonistic idea, why should he deprive himself of his direct, immediate pleasure? and how could we expect in him, an uneducated child, the formation of the far greater and more general conceptions of Property, Right, Justice, and Punishment?

These examples of the psychological factors involved in the acts of these youthful delinquents put us in doubt, from a new point of view, as to our right to punish the offenders; but, if one considers their psychology as a whole, beyond the question of Right, one eliminates even the criterion of the utility and suitability of such punishment. Let us then see who these youthful criminals are, and where they come from.

I do not wish, nor could I, speak of the youths who become criminals because of pathological causes, either personal or hereditary (because of epilepsy, imbecility, constitutional amorality, posthumous effects of encephalitis lethargica, etc.), which come in the medical field (A. Ley).

The other delinquents can be called normal, it being normal for children in their conditions to become such, especially if they are intelligent and have varied interests, and especially if they possess a

certain individuality, but are lacking the corrective and educational foundation of a well-constituted family.

It is an old idea that youthful criminality is the most direct result of the dissolution of the family nucleus. This dissolution at first seemed allied to industrial progress. But the War, with its typical consequences of children being left to themselves without any authority or curb, and the troubled times after the War which complicated these consequences still more, especially by the increase in the cost of living and the greater insufficiency of lodgings, all these have reinforced these reasons and caused the increase in delinquency which is today a problem in all nations.

If we think of the kind of life that a poor child of ten or twelve must lead, in the single room where usually he must live with his parents, his brothers and sisters, badly fed, always in the same way, dressed as best may be, without any possibility of control, of assistance, or of encouragement towards good or improvement, it is only too evident that anything different, more pleasant, that may come his way, must seem to him very desirable, and that the penalty inflicted by Society—that is, prison—cannot be to him anything worse than his present conditions. It is quite natural then that the child should try to drift away from his family, to try and live better, eating that which he finds, sleeping wherever he can find temporary refuge.

Of course, to us, for whom the table is laid twice or three times a day, who have a good bed, and every convenience, newspapers, the post, etc., the conditions in which run-away children find themselves seem absolutely frightful. But to the greater number of children who tend to become delinquents they seem almost ideal. They are upheld, illuminated by their faith in liberty, which is, indeed, the mirage that makes so many martyrs!

Apart from all comparative valuation, it appears quite evident that for a lad who is physically well-constituted, no profession could be more suitable and attractive than that of the criminal, because *it satisfies his tendencies and instincts as soon as they appear and flourish.*

Lesser and greater criminality, in fact, offer to children the physically exuberant life, the possibility of adventure, the unforeseen, the abolition of *routine*, of the tiresome systematic occupations, of the continual grumbling, always tedious and often uncalled for on the

part of parents, instead of satisfying their greed, and often with important material gain. It is the satisfying of all these tendencies to which our own children give vent, as we have seen, by means of games, matches, sports, reading books of travel and adventure, going for trips, visiting exhibitions, the theater, moving pictures, etc.

Not only this, but little by little the adventurous life of the youth takes on a particular psychological tinge, assuming *an aesthetic character*. This is a most important element, because it takes root, and confirms those tendencies in the individual, urging him to improve and perfect them. Furthermore, the aesthetic character gives to his profession an immanent, general interest, for it is a particularity of this interest to expand into similar fields. Finally, it makes these tendencies become incurable, for a more important reason, one which deserves a more ample explanation.

As a rule, when it is not a question of children who are absolutely deprived of relatives or have been forsaken by them, those who drift towards occasional criminality (theft and prostitution) are individuals who, for intimate or outside reasons, have not succeeded in adapting themselves to the surroundings in which they were born and have been growing up. This deviation is felt by them as something which is incomplete, inasmuch as it corresponds to a lacking of personal synthesis. This synthesis, however, can always be formed on a different level, as we might say, and it is criminality which offers this new level. It is then that, aided by circumstances, the youth is no more, in his own idea, one who steals, but he has become "*the thief*"; and the young girls who have not yet reached puberty, but who sell themselves out of complaisance or for money, wish to be "prostitutes," to take up that calling, even if they can only exercise it with boys or, occasionally, with degenerates.

The magistrates then say that the "occasional criminal" has become a "habitual criminal," but the root of the question is always the same: that the child has *been obliged* to form a new personality, which is a criminal personality. Now, this second personality, having been created by himself, corresponds to his innermost needs and longings, and it is in its application that the child or youth finds his greatest satisfaction.

A further proof of the importance of this explanation is to be found in the fact of the different "categories" of criminals, and, as a correlative phenomenon, one can note the sentimental, almost or-

ganic incapacity, for instance, towards theft, in certain youths at first educated, and then finding themselves "déclassés" owing to misfortune, and who almost become idiosyncratic, that is, they fear terribly that they might end by stealing.

To conclude, therefore, it appears that the youths who have not become criminals owing to any disease, but who, on the contrary, are intellectually normal, have for the most part been urged towards criminality by the abandonment to which they are left by their families, and by the lack of education of any kind.

Criminality arises in them from mental habits which form themselves naturally in children, and which are favored and enhanced by their surroundings. These tendencies need only the slightest encouragement to let the children obtain all that which they wish for, and they have full opportunity to develop and to become firmly established, until they end by constituting the true personality of the child. These tendencies, in fact, become successively habits, then impulsions, because interest makes them easier and then necessary, and, finally, because they assume for the child a certain aesthetic coloring.

Naturally, normal persons cannot understand and therefore evaluate the mental make-up of these children unless they are able to forget all that which forms the basis of their moral sense and of their habits of social behavior. Because the conception of moral duty has become so natural to the greater part of society, it is with difficulty that an individual can succeed in forgetting that it is not a natural thing, but, instead, is the result of education.

For the foregoing, it is easy to deduce what ought to be the course of the therapeutic treatment of youthful criminality.

If we admit that I have seen aright during the thirty years in which I have studied the question, and that things really stand as I have tried to describe them to you, all of us, I think, will fully agree that that which is offered as a treatment for these children by our laws, i.e., imprisonment, is perfectly useless, as is also our system of reformatories.

Prison may frighten until one has come to know it. But for experienced children it is merely a place where they are better off than at home, where they find many satisfactions (often even sexual), and where they form most useful acquaintances for their future criminal careers.

Furthermore, and this is even worse, the children soon get accustomed to considering their punishment as a professional risk, and of the same value, though in contrast to it, as their own acts. Therefore, having gone through their term of punishment, they consider themselves purified, and to be "quits" with everyone. And, just as they will not accept rebuke for the fault of which they have been freed by the punishment endured, so also they no more feel the weight of that which they have done.

In Italy, besides the prisons, the government maintains different reformatories for delinquent children, one for those who have already received sentences, the others for children who are only considered "naughty." These institutions ought to supplement the training which the children should obtain at home. However, since they care for a great many unselected children and classify them, as a rule, by their height or by age, instead of according to their tendencies and problems, these institutions probably do not give all the results desirable, in spite of intelligent and capable directors and the backing given them by the government.

Better results have come from some private institutions. Here and there several colonies have grown up in which these children have been cared for and trained. After the War, some orphan homes adopted special means of caring for these "outlaws." But these have assumed only very limited proportions. The outstanding one is the great Colony of Arese, near Milan, which is backed by a number of wealthy philanthropists. Of all the interesting things about this colony the most interesting to me is that all of its efficiency, order, and results depend entirely upon the personality of those who make up its staff. Force is considered the enemy of discipline, which, to be efficacious in the formation of character, and for the education of the youth, must be *accepted freely*. The child must feel that discipline is in no way inflicted in the interest of the person giving the order. He should be able to see that it is really useful to himself to obey. Then he is likely to adapt himself and be willing to do as he has been told.

I have left to the last an institution which is intermediate between those of the state and the privately conducted ones because it is the best and the most characteristic type that Italy can present, even if, during the last few years, it has not maintained its earlier standards.

About thirty years ago, an old sea captain, a certain Garaventa, gathered around him on an old ship which was in disuse and was anchored in the port of Genoa all the lost or abandoned boys he found wandering about the streets. The idea was approved by the State after ten or fifteen years, and Garaventa received some financial support. Later on the Government turned over for the same purpose another antiquated vessel, the "Caracciolo," in the port of Naples, and still another, the "Scilla," in Venice.

A visit to one of these ships impresses upon one the importance of the personality of those who are in charge. Two of the ships are in charge of retired naval officers, while the educational program of the third is under the direction of Signora Giulia Civita Franceschi, who has had remarkable success in dealing with the boys who have come under her influence. Her unusual success leads me to believe that it is possible, and I might even say easy, to set right the perverse tendencies of delinquents who are normally intelligent, so long as one possesses the necessary disposition to know "how to take" them. As a matter of fact, not all of these unfortunate youths "succeed" in becoming criminals, and those who do are those who are more intelligent and have more varied and changeable interests. It is for this reason that, once they are set on the right path, they often give the best results.

One of my most interesting experiences had to do with an experiment which I undertook, from 1910 to 1915, at Imola in the two free colonies for "youths of abnormal character" (as youthful criminals might much more suitably be called). The magistrate who was then President of the Law Courts of Bologna sent to me from time to time about twenty children of both sexes who would otherwise have been sent to a reformatory or who had just finished their term of imprisonment. I asked Professor Gabriella Francia, who was then working with me as a special teacher in the section for deficient of my lunatic asylum, to take this group of delinquents, along with 18 or 20 mentally deficient children, to an isolated villa that had been offered to me. As assistants, I could only offer her an old nurse and four quiet patients, specialized in particular forms of work, such as in making bread, washing, mending, or keeping accounts. My idea was that the children should live more or less in the modest type of surroundings to which they had been accustomed before being sent to the asylum, and to which they would have to return if they were released. Miss Francia has told of this experiment in an article in

the *Revista di psicologia* (1910). Unfortunately, the Government was forced to withdraw their very slight financial contribution to our colony because of war expenses, so our work was discontinued.

The general principles upon which we worked were as follows: All the children (intellectually deficient or abnormal) were, *on principle*, treated in the same way; that is to say, they were requested to occupy themselves, since "work is a normal condition of healthy life." If a child did not want to work, a thing which soon became exceptional, a suitable reason or excuse for his doing it was sought and suggested to the child. Usually, if no physical cause prevented him from working, the child would turn to some other work, which he found more pleasant or more suitable to his inclinations at that moment. But very often it was noticed that he would turn back to the first task; having saved his point of honor by showing himself undisciplined, or at least not wanting to accept advice, he would very soon, if not at once, understand that the work had been suggested with a definite aim, the teacher considering it the most suitable for him. It must, of course, be understood that the teacher always had to choose work in which it was probable that the child should succeed so that he could find satisfaction in doing it. What is even more important, however, is to make the child himself notice the advantage which comes to his own personality through his work, the problem being to lead the child on, by means of patient, hidden, direct or indirect suggestion, to apparently make this discovery for himself.

In this way indirect moral education was brought about without any moral preaching. This type of treatment is unusual especially because the teaching must be carried out without ever being seen or openly felt by the children.

The method presents several advantages, all of them contributing toward the modification of the primitive manner of living practiced by the children and toward a more suitable and moral adjustment to society. In the first place, placing the delinquents in an institution with children who are *evidently* and seriously intellectually deficient gives them the impression that they, too, are considered abnormal, that their acts are considered the result of their morbid, hence inferior, condition (an excellent indirect lesson in modesty).

The second advantage is ethical. The criminal youths find that they are treated in the same way as those who are intellectually deficient, that their criminal acts or their idleness are not punished. They find that they are merely advised and corrected, and they come

to feel the influence of the society that has discarded them and that possesses an intangible force which now, for the first time, does not display itself in the form of vengeance, but in the form of an attempt—which they may deride, but which they still appreciate—at their *education*.

As a third advantage, I shall mention the urging towards self-education. They see the same acts evaluated in the same way, whether committed by themselves or by those who are intellectually deficient. They therefore have apparent proof that their intelligence is not sufficient to set them above their more unfortunate companion in the good graces or the affection of their teacher, whom, little by little, they come to regard as the representative of the whole outside world and of its authority. In order to attain favorable attention all their efforts must be bent in the direction of social conformity, and they fully understand that all their victories in that field will be fully appreciated.

And, finally, I shall mention another ethical advantage. Living on the same footing with children who are intellectually subnormal gives the normal children a sympathetic feeling almost from the very first day. This seems to come about by means of a mechanism which tends towards self-exaltation, in *comparison with those unfortunate creatures*, and this new sense rapidly develops into *pity*, which is the primordial element of social intercourse.

One other favorable outcome of our experiment lay in the fact that it gave all those who observed it *absolute faith* in its practical possibilities. This faith had already been the possession of America, where the first reformatories were instituted. I do not know whether statistics prove that these reformatories have had a practical efficacy in the sense that they have prevented delinquents from becoming adult criminals. But I can say, on the basis of my own experience, that children who are said to be “hopelessly lost” are only misled and that they are called “lost” simply because no one has ever taken the trouble to look after them; that all that is necessary is to direct their interest and their natural tendencies toward useful ends and they will gladly become once more useful members of society. I would also say that the most important factor in accomplishing this reconstruction of youth is the personality of the one who works with the children. Kindness and intelligent faith are the essential requirements for success on the part of such a person.

LE PROBLÈME DE LA CRIMINALITÉ CHEZ LES ENFANTS

(Résumé)

On montre que la soi-disant criminalité ou délit chez les enfants n'est que le résultat de certaines tendances naturelles qui n'ont pas été bien dirigées à cause d'un manque de conditions favorables à la maison ou à cause d'un manque entier de vie familiale. Il paraît que le meilleur remède est de donner à ces enfants l'occasion de faire sortir les tendances naturelles sous la direction, pleine de tact, d'un certain type de personnalité dont la compréhension, la bonté, et une foi intelligente sont les qualités essentielles.

FERRARI

DAS PROBLEM DER KRIMINALITÄT BEI KINDERN

(Referat)

Es wird erwiesen, dass so-gennante Kriminalität oder Verbrechen bei Kindern nur die Auswirkung gewisser natürlicher Neigungen ist, die infolge der Abwesenheit günstiger häuslicher Verhältnisse oder infolge völligen Mangels an Familienleben nicht die richtige Lenkung gefunden haben. Es zeigt sich, dass die beste Behandlungsweise darin besteht, dass man diesen Kindern die Gelegenheit bietet, natürlichen Neigungen Ausdruck zu geben, unter der taktvollen Lenkung eines gewissen Persönlichkeitstypus bei dem Verständniss, Güte, und intelligentes Vertrauen (faith) wesentliche Eigenschaften darstellen.

FERRARI

A SCALE FOR MEASURING PERSISTENCE*¹

From the Psychological Laboratories of the University of Chicago

CHARLES K. A. WANG

The present form was compiled for the purpose of providing a scale for measuring persistence, using the term as it is ordinarily understood. We are not concerned with the nature of persistence but wish merely to determine whether there is any difference between the behavior of individuals who are considered persistent and that of those who are not so considered.

The scale consists of 111 questions each of which has one answer that has been determined to be diagnostic of positive persistence. So that it may serve its purpose best by making the name noncommittal, the scale is called "A Self-Appraisal Schedule."

JUDGMENT OF PERSISTENCE

The preliminary form of the schedule consisted of 181 questions. These were collected and developed from a large number of sources, including individual descriptions of persistent behavior, Woodworth's Psychoneurotic Inventory, Allports' ascendance-submission test, and Thurstones' Personality Schedule.

The first step in our procedure was to submit the questionnaire to 75 individuals (college students, teachers, graduate students, professional men, etc.) who acted as judges. Each of the judges was asked to indicate which, if any, answer to every question he would consider as typically descriptive of the behavior of a persistent person.

After tabulating the frequencies of the judges' responses to each of the questions, it was found that, while there was poor agreement on some of the questions, there was almost perfect agreement that particular answers to many other questions indicated positive persistence. The situation called for an arbitrary basis of selection, and the one decided on was as follows: First, no question would be considered for tentative scoring if more than 40% of the judges failed

*Accepted for publication by L. L. Thurstone of the Editorial Board and received in the Editorial Office, February 24, 1931.

¹The writer is indebted to Dr. F. A. Kingsbury for his interest and suggestions and to Dr. L. L. Thurstone for data on intelligence and neurotic scores.

to vote on it. Secondly, if only 60% of the judges voted on a given question, there must be no disagreement whatever. Thirdly, if more than 60% of the judges voted on a given question, the difference between the "yes" vote and the "no" vote must be greater than 60% of the total number of judges. To illustrate, if 70% of all the judges voted that "yes" response to a given question constituted persistence, there must be no more than 10% of them who voted the "no" response.

Sixty questions out of the 181 met the above specifications and were therefore used in the tentative scoring of the subjects' papers.

TENTATIVE SCORING

The same questionnaire of 181 questions, but with a different set of instructions, was given to 512 undergraduate students in the University of Chicago. Of this number, 97 were seniors, 198 sophomores and juniors, and 217 freshmen. Classified according to sex, 299 were men and 213 were women. The subjects were simply asked to answer the questions according to their own experience by drawing a ring around "yes" or "no" before each question. At the completion of the questionnaire, each subject was also asked to rate himself on persistence, using a seven-point scale.

The frequency distributions of the tentative scores are presented in Table 1. It will be observed that while the senior and junior-sopho-

TABLE 1
FREQUENCY DISTRIBUTIONS OF TENTATIVE SCORES ON THE SELF-APPRAISAL SCHEDULE

Score	Seniors	Jr.-Soph.	Freshmen	Total
53-56	2	8	17	27
49-52	16	23	31	70
45-48	13	24	51	88
41-44	15	42	43	100
37-40	13	36	24	73
33-36	16	22	22	60
29-32	11	12	12	35
25-28	8	13	6	27
21-24	0	11	5	16
17-20	2	3	4	9
13-16	0	2	2	4
9-12	0	1		1
5-8	1	1		2
Total	97	198	217	512
Mean	38.87	38.90	41.95	40.27

more groups are very similar, the freshmen group is notably higher in scores. Upon inquiry in the University Examiner's Office, it was revealed that the freshmen group might be superior for two reasons: first, the 1930 freshmen of the University of Chicago were selected from a much larger number of applicants for admission to the University than in any previous year; secondly, the particular group of freshmen who filled in the questionnaire belonged to the General Survey Class, admission to which was by selection and invitation. This detection of superiority among the freshmen at the stage of tentative scoring came as a surprise but not without gratification.

SELECTION OF QUESTIONS FOR THE REVISED SCORING

On the basis of the tentative scores, two groups were segregated: the highest 10% in scores constituting Group I, and the lowest 10% in scores constituting Group II. Then the responses to each question by every individual in these two groups were tabulated. The retention or exclusion of a question from the revised scoring was then determined in each case by the ratio of the number of "Yes" answers in Group I to be the corresponding number in Group II, considered in conjunction with the like ratio of the "no" answers. And the relative frequencies of these answers determined whether the "yes" or the "no" response was to constitute a score for positive persistence. This is an application of the criterion of internal consistency.²

In the appendix are presented the list of questions retained for the revised scoring, the answers which count for positive persistence, and the frequencies of response by Groups I and II. In order to save space, the questions found less diagnostic have been omitted from this list.

FREQUENCY DISTRIBUTIONS OF REVISED SCORES

After all the papers have been re-scored in accordance with the list shown in the appendix, it was found that the variation in scores was from 9 to 109. In Table 2 are presented the frequency distributions of total scores on the Self-Appraisal Schedule for the 512 undergraduate students. The superiority of the freshmen group over the upper-class groups appears again, this time even more notably than in the tentative scores. The difference between the mean of the freshmen and that of the seniors is 7.26 with a *P.E.* of 1.38, while the difference between the mean of the freshmen and that of the

²See (1).

TABLE 2
FREQUENCY DISTRIBUTIONS OF SCORES ON THE SELF-APPRAISAL SCHEDULE

Score	Seniors		Tot.	Jr.-Soph.		Tot.	Freshmen		Tot.	Men	Total	
	Men	Wom.		Men	Wom.		Men	Wom.			Men	Wom.
105				1		1	5		5	6		6
100				5		5	7		10	14		17
95	2		2	6	1	7	15	6	21	23	3	31
90	2	1	3	10	1	11	12	7	19	27	10	37
85	5	2	7	4	11	15	15	11	26	22	27	49
	3	5	8									
80	7	4	11	22	6	28	20	13	33	49	23	72
75	4	10	14	11	9	20	18	8	26	33	27	60
70	1	6	7	7	9	16	13	6	19	21	21	42
65	6	6	12	16	12	28	10	7	17	32	25	57
60	4	4	8	8	13	21	3	4	7	15	21	36
55	5	5	10	9	3	12	7	6	13	21	14	35
50	3	2	5	8	4	12	0	6	6	11	12	23
45	3	1	4	2	3	5	2	2	4	7	6	13
40	2	1	3	1	3	4	2	1	3	5	5	10
35	0	1	1	3	5	8	1	2	3	4	8	12
30	1		1	1	1	2	2	1	3	4	2	6
25	0		0	1	0	1	1		1	2	0	2
20	0		0	0	0	0	1		1	1	0	1
15	0		0	0	1	1				0	1	1
10	0		0	0		0				0	0	0
5	1		1	1		1				2		2
Total	49	48	97	116	82	198	134	83	217	299	213	512
Mean	70.0	71.1	70.5	72.6	66.7	69.5	79.5	74.9	77.8	75.2	70.1	73.5

junior-sophomore group is 8.26 with a *P.E.* of 1.14. The statistical significance of these differences is therefore unquestionable.

VALIDITY

The validity of the Self-Appraisal Schedule as a measure of persistence has been considered from two approaches. It will be recalled that a self-rating was obtained from each subject. The correlation coefficient between this self-rating and the revised score on the Schedule was found to be $+.522$, which is higher than one might expect for this sort of correlation. The more important consideration in establishing the validity of the Schedule, however, is the application of the criterion of internal consistency. The selection of questions on the basis of the frequencies of response by the high- and low-persistent groups shows that the manner in which the answers were scored was at least consistent. Thurstone regards this consistency as more essential in establishing validity than correlations with outside criteria even when they are available. This consistency also proves that there is a common core of some kind throughout the questions that were finally retained, and, since we have obtained very close agreement among the judges that particular answers to the questions describe persistence in the first place, we are thus justified in regarding the Schedule as valid for measuring persistence.

RELIABILITY

The reliability of the total score has been estimated by correlating the scores in parallel halves of the Schedule. In the original printing, the questions were distributed on eight mimeographed pages. The correlation between scores on the odd- and even-numbered pages was found to be $+.84$ and the estimated reliability coefficient for the whole Schedule is therefore $.912$. This is, in a way, corroborated by a correlation coefficient of $+.921$ obtained between the tentative scores based upon 60 questions and the revised scores after applying the criterion of internal consistency.

CORRELATION WITH INTELLIGENCE AND NEUROTIC DISPOSITION

During the last three years all the entering freshmen at the University of Chicago have been required, in the Freshmen Week, to take an intelligence examination and to fill in the Thurstones' Personality Schedule. Since, then, with the exception of the seniors, most of our subjects have their intelligence scores and scores on the Personality Schedule on file, it was relatively easy to ascertain whether

any relation exists between persistence and intelligence and between persistence and neurotic disposition. The correlation between scores on the intelligence test of the American Council of Education and the scores on the Self-Appraisal Schedule was found to be -0.05 , with a *P.E.* of $.04$, indicating that there is practically nothing in common between persistence and intelligence, at least as measured by the scales in question.

In correlating scores on the Personality Schedule and the Self-Appraisal Schedule, a coefficient of $+.51$ was obtained. This seems high, but it is not surprising. While lack of persistence is not necessarily a factor of maladjustment, it is reasonable to suppose that a "quitter" is not a well-adjusted individual and that one who shows persistence in his adaptive behavior is, in doing so, exhibiting adjustment.

APPENDIX

THE SELF-APPRAISAL SCHEDULE AND A COMPARISON OF RESPONSES BY HIGH- AND LOW-PERSISTENT GROUPS

Question	Pers. Ans.	High		Low	
		Yes	No	Yes	No.
Can you endure strong physical pain?	Yes	44	5	21	25
Do you usually feel sure that you are right in your beliefs?	Yes	46	6	28	21
In a difficult problem, is it genuinely painful for you to give up?	Yes	45	4	7	41
Are you easily discouraged?	No	0	50	37	10
In an argument, do you find it difficult to give in?	Yes	45	5	21	28
Are you considered "steady" and not "flighty"?	Yes	47	0	25	23
In the organizations that you belong to, are you usually satisfied with being merely a member?	No	5	44	22	24
If you see someone that you desire to know, do you often make repeated efforts to effect a meeting or introduction?	Yes	39	11	13	34
Do you find it difficult to refuse a solicitor for funds if you are not interested in the cause?	No	4	46	34	13
In keeping a social or business engagement, are you usually prompt?	Yes	46	4	18	30
Do you usually make a schematic plan in gaining your ends?	Yes	43	6	18	30

Question	Pers. Ans.	High		Low	
		Yes	No	Yes	No
Are you easily influenced by convention or public opinion?	No	7	42	29	19
Do you get over depressive moods easily?	Yes	42	8	22	18
When things go wrong can you take it calmly?	Yes	47	2	27	20
In an argument do you try to meet every point with a counterpoint?	Yes	48	2	33	17
When you are confronted with a difficult problem, would you give up after one or two trials?	No	0	50	37	10
In selling or soliciting do you accept "no" readily and try the next prospect?	No	3	42	37	8
Do you often day-dream?	No	24	26	44	5
Are you usually insistent in getting exactly what you want?	Yes	45	5	19	29
In a competitive game, are you encouraged to play harder if your opponent is reputed to be a better player than yourself?	Yes	50	0	38	10
Do you consider that your interests are rather narrow in range?	No	3	47	18	31
Do failures have long depressive effects on you?	No	5	44	17	30
Do you usually stay by a task until it is finished?	Yes	49	1	14	32
Are you willing to sacrifice comforts for future gains?	Yes	49	0	26	20
Do you hold your opinions firmly?	Yes	47	3	24	22
Are you rather easily upset?	No	2	48	23	25
Do you have difficulty in concentrating your thoughts?	No	7	43	40	5
Do you frequently build air castles?	No	25	25	42	5
In performing a planned task are you guided by other people's remarks or criticisms?	No	26	23	41	5
Can you give good reasons for your actions?	Yes	49	0	32	16
At a reception or tea do you seek to meet the important person present?	Yes	31	18	10	36
In purchasing articles, do you frequently accept substitute brands for one of your preference?	No	2	48	19	29
When you see in a large crowd someone to whom you wish to speak, do you insist on reaching him even if it involves considerable trouble?	Yes	37	13	6	41

Question	Pers. Ans.	High		Low	
		Yes	No	Yes	No
Are you sensitive to other's opinions of your acts?	No	22	26	44	5
Have you ever had more than transient periods of depression or elation?	No	5	42	15	32
In case your corner news-stand ran out of the paper you want, do you frequently go to another stand, perhaps some distance away, to get it?	Yes	41	8	22	21
Do you borrow money for things that are not urgent needs?	No	1	49	14	36
Do you frequently yield to pleasure at the neglect of duty?	No	3	46	40	8
Are you considered resourceful?	Yes	45	3	23	19
Do you make repeated efforts to recover a lost article?	Yes	43	7	12	31
Can you work on tasks that involve tiresome routine over an extended period of time?	Yes	47	3	13	36
Are you usually patient?	Yes	40	10	23	23
Do you usually control your temper?	Yes	45	4	26	13
Do people regard you as extremely systematic?	Yes	30	15	4	43
Do you lose your head easily in a complicated situation?	No	1	49	15	31
Do you worry too long over humiliating experiences?	No	8	42	33	16
Are you sometimes the leader at a social affair?	Yes	41	9	16	33
As a rule do you prefer to work out a thing for yourself?	Yes	48	2	26	24
Do you have the habit of leaving tasks unfinished?	No	0	45	35	15
Are you usually cool and composed in a dangerous situation?	Yes	47	3	33	13
Are you enthusiastic about your life's possibilities?	Yes	47	3	21	20
Do you get discouraged easily?	No	0	50	39	8
Do you often say things on the spur of the moment and later regret it?	No	21	28	41	8
Do your interests change quickly?	No	2	48	28	18
Are you regarded as particular about details?	Yes	37	12	12	34
Do you allow people to crowd ahead in line?	No	3	47	32	15

Question	Pers. Ans.	High		Low	
		Yes	No	Yes	No
Can you stand criticism without feeling hurt?	Yes	42	7	17	27
Do you like to take on responsibilities?	Yes	45	5	10	37
Do you work better when you have to convince people of your ability?	Yes	41	9	26	21
Do you dislike selling things?	No	18	32	36	13
Does your mind often wander badly so that you lose track of what you are doing?	No	6	44	29	20
Are you frequently absent-minded?	No	6	44	24	24
Do you feel tired most of the time?	No	1	49	12	37
Do you ever volunteer to differ with a teacher and argue the point out?	Yes	47	2	30	18
In public speaking do you repeat your points until you feel sure that your audience has fully understood you?	Yes	39	11	11	38
Do you frequently have the feeling that life is futile?	No	7	43	23	23
Do you usually stick to a task until it is completed?	Yes	50	0	13	35
In working out a problem do you usually limit yourself to that task to the exclusion of others?	Yes	30	20	8	42
Having determined to do or to obtain something do you usually stick to it until you have attained your goal?	Yes	50	0	17	29
Do you consider yourself tenacious?	Yes	42	6	6	38
Are you easily persuaded to change your opinion or belief?	No	1	48	34	13
Do you indulge in self-pity when things go wrong?	No	4	46	25	22
Do you get rattled easily?	No	4	46	15	31
Are you strongly motivated by ambition?	Yes	43	7	16	30
Do you dislike and avoid any process of selling?	No	13	37	32	15
Do you prefer to work things out on your own hook?	Yes	49	0	25	23
Do you shrink from facing a crisis?	No	2	48	24	22
Do you work by fits and starts?	No	0	50	40	7
Do you like work which requires painstaking manipulations?	Yes	34	16	6	42
Do you persevere in spite of failure?	Yes	50	0	13	33
Do you approach difficult tasks methodically?	Yes	48	2	12	35

Question	Pers. Ans.	High		Low	
		Yes	No	Yes	No
Do you work according to some schedule or plan?	Yes	45	4	10	37
Are you extremely energetic?	Yes	37	13	7	41
In a difficult task do you give up rather easily?	No	0	49	35	10
Can you say "no" to a salesman who has taken pains in showing you merchandise which does not entirely suit you?	Yes	46	4	23	23
Do you approach difficult tasks resolutely?	Yes	50	0	20	25
In a social gathering do you frequently take the initiative to enliven the party when it becomes dull?	Yes	41	9	20	29
When you are served an inferior dish at a high class restaurant do you complain about it to the waiter?	Yes	23	26	8	41
In conversation do you frequently insist on presenting your point when others have changed the subject?	Yes	23	27	9	39
Are you shy with girls (or boys)?	No	5	42	15	30
Does criticism disturb you badly?	No	5	45	26	22
Do you think that you lack self-confidence?	No	4	44	32	16
Do you like to solve puzzles?	Yes	39	11	19	30
Do you usually plan your work ahead?	Yes	49	1	13	37
Do you "hang on" to your opinion or belief?	Yes	46	3	19	29
Do you let yourself go when angry?	No	3	46	16	33
In general, are you confident about your abilities?	Yes	46	4	20	28
Are you troubled with dreams about your work?	No	5	45	15	32
In reflective moments, are your thoughts more often about future events or conquests than about past possibilities?	Yes	45	4	33	16
At a high class restaurant do you frequently verify the bill if you find it slightly larger than you expected it to be?	Yes	38	12	8	41
Do you usually know just what you want to do next?	Yes	47	3	12	36
Do you have trouble in making up your mind?	No	6	44	35	11
Are you afraid of responsibility?	No	1	49	23	23

Question	Pers. Ans.	High		Low	
		Yes	No	Yes	No
Do you get tired of work quickly?	No	0	50	39	9
Do you frequently make plans for your future work?	Yes	48	2	20	28
In working out a problem do you feel a strong tendency to finish it in spite of fatigue or other distraction?	Yes	50	0	17	31
On resuming an unfinished task after an interruption, do your former thoughts on the subject come back readily?	Yes	44	6	27	23
When unexpectedly asked a question which you know well enough but have not been thinking of at the time, can you answer readily?	Yes	35	13	15	35
In a difficult or distasteful task, are you easily diverted?	No	2	48	44	6
After you have made a decision, are you easily influenced by diverse suggestions?	No	2	48	33	15
In action or speech can you always defend your position rationally?	Yes	45	5	23	26

REFERENCE

1. THURSTONE, L. L., & THURSTONE, T. G. A neurotic inventory. *J. Soc. Psychol.*, 1930, 1, 3-30.

University of Chicago
Chicago, Illinois

UNE ÉCHELLE POUR LA MESURE DE LA PERSISTANCE

(Résumé)

Le but de cette échelle est de nous donner une assez constante mesure de la persistance, employant le terme dans son sens usuel. L'échelle se compose de 111 questions dont on peut répondre à chacune "oui" ou "non", et l'une ou l'autre réponse est diagnostique de la persistance positive. La première partie du processus expérimental a été de soumettre une plus longue forme préliminaire de l'échelle à 75 juges à qui on a demandé d'indiquer si une réponse, et laquelle, pourrait être considérée typiquement descriptive du comportement d'une personne persistante. Les réponses des juges ont donné une base pour l'évaluation tentative des questionnaires des sujets. La seconde partie du processus a été de donner l'échelle à 512 étudiants universitaires ("undergraduates") et d'obtenir de chacun d'entre eux une évaluation de soi à l'égard de la persistance. On a établi la valeur de l'échelle, premièrement, par une corrélation de 0,52 obtenue entre les évaluations de soi à l'égard de la persistance et les résultats sur l'échelle;

et deuxièmement, par l'application du critère de constance intérieure aux résultats tentatifs qui ont été basés sur l'accord des juges sur le comportement persistant. On estime la constance de l'échelle 0,92 pour les 111 questions. On a obtenu une corrélation de +0,51 entre cette échelle et l'inventaire névrosique de Thurstone, mais la corrélation avec un test d'intelligence a été presque nulle.

WANG

EIN MASSSTAB ZUR MESSUNG DER BEHARRLICHKEIT

(Referat)

Dieser Massstab hat den Zweck, als ziemlich zuverlässiges Mass der Beharrlichkeit (persistence), im gewöhnlichen Sinne der Wortes, zu dienen. Der Massstab besteht aus 111 Fragen, von denen jede mit 'ja' oder 'nein' beantwortet werden kann, wobei die eine oder die andere Antwort die Anwesenheit von positiver Beharrlichkeit erweist. Der erste Teil des experimentellen Verfahrens bestand darin, dass man eine vorläufige und längere Form des Massstabes 75 "Richtern" vorlegte, die anzudeuten hatten, welche, wenn überhaupt eine, Antwort auf jede Frage als typisch beschreibend für das Verhalten einer beharrlichen (persistent) Person gelten würde. Die Antworten der Richter dienten als Basis der vorläufigen Zensierung (scoring) der Papiere der Versuchspersonen. Der zweite Teil des Verfahrens bestand darin, dass man den Massstab 512 Studenten vorlegte und von jedem dieser Studenten eine Selbstabschätzung (self rating) in Bezug auf Beharrlichkeit erhielt. Die Gültigkeit (validity) des Massstabes wird festgesetzt, erstens durch einen Korrelationswert von .52, der zwischen den Selbstabschätzungen der Beharrlichkeit und den an dem Massstab erhaltenen Zahlen bestand, und zweitens durch die Anwendung des Kriteriums der innerlichen Konsequenz (internal consistency) auf die vorläufigen Zahlen, die auf die Uebereinstimmung der Richter in Bezug auf beharrliches Verhalten (persistent behavior) ruhten. Die Zuverlässigkeit des Massstabes wird auf .92 für die 111 Fragen abgeschätzt. Es ist ein Korrelationswert von +.51 erhalten worden zwischen diesem Massstabe und dem Inventarium der Nervosität, von Thurstone (neurotic Inventory). Mit einer Prüfung der Intelligenz war die Korrelation aber fast Null.

WANG

SHORT ARTICLES AND NOTES

THE SOCIAL ADJUSTMENT OF DELINQUENTS WHO ARE UNABLE TO INHIBIT OLD AUTOMATIC PERCEPTUAL RESPONSES¹

JAMES QUINTER HOLSOPPLE

The observations presented in this paper comprise one part of an experimental psychological approach to the general problem of criminal recidivism. Any psychologist whose human material consists of convicted and committed thieves, murderers, embezzlers, and prostitutes, while learning the inadequacy of the old superstition that criminals always return to the scenes of their crimes, should be more impressed than most psychologists usually are with a heavy obligation to explain the scandalously high frequency with which criminals do return to the scenes of their "reformations." By scientific or quasi-scientific studies, various persons have persistently tried, and as persistently failed, to discover that pathological germ in the mental makeup of the recidivist which prevents him from understanding that the road to his happiness lies in the direction of social conformity. In New Jersey every admission to a state institution is examined in different ways, psychological among others, to the end that a course of living, work, and instruction shall be determined which may enable him subsequently to remain outside of the institution and out of serious conflict with the current mores. Our constant search for more adequate examination methods has resulted in our use of *mirror drawing*, an old technique, but one which seems to have been almost completely lost in the shuffle of mental testing. The mirror-drawing performance is difficult for individuals are unable to inhibit those old automatic perceptual responses which are involved in drawing a line from one visually observed point to another. The results of our work show that this difficulty is reflected in the general behavior of those delinquents whose apparent inflexibility of habit seems to prevent their unlearning those types of misbehavior which repeatedly get them into trouble. Our data also indicate that those recidivists who seem to be rather the victims of an overwhelmingly unfavorable environment than to have deep-seated personal disabilities show a minimum of difficulty in their mirror drawing. The value of early recognition of those delinquents with a high capacity for recidivism is too obvious to need elaboration.

Until recently the localization of the roots of recidivism in mental deficiency has been in fairly good repute. Hence it must be pointed out that

¹Miss E. P. Connors, Miss Gertrude L. Corbett, and Mr. Donald T. Griffin deserve particular mention for their efforts in collecting the material upon which this study is based.

in no sense is mirror drawing here submitted as a substitute for measures of general mental level or as a new "intelligence" test. Whether or not criminal recidivists are, as a whole, more or less intelligent than the general population is not now, and never should have been, a time-consuming problem. Criminal recidivists are not "as a whole," they are individuals. A few are highly intelligent, based upon any definition other than one involving the recidivism itself, and many are clearly feeble-minded. Almost any scientific program, practical in application and humanitarian in motive, demands that the seriously deficient recidivist be treated in view of his feeble-mindedness. The real *problem* of recidivism lies in the clear majority of recidivists who can, under no circumstances, be labeled feeble-minded. Repeated successes of parolees who are really too stupid to stay out of trouble, combined with repeated failures of parolees whose IQ's are high enough to enable them to stay out of trouble, point clearly enough to the hypothesis that of the mental factors involved in recidivism, general mental level does not exert the major influence.

The selection of mirror drawing as a situation which might be scrutinized for indications of mental characteristics important in recidivism was based upon the following observations and hypotheses:

1. In the entire literature of tests and experiments, insofar as they have been applied to delinquents, there are no data more interesting, and few data *as* interesting as those obtained from the mirror-drawing procedure.

2. Time after time, a returned parole violator, whose intentions upon release seemed to be of the best sort, and who might be justified in explaining away his violation, has observed with apparent candor that there just seemed to be no reason for his lapse except that he simply could not avoid doing as he always had.

3. The mirror-drawing procedure seemed to be relatively easy of analysis, successful performance depending largely upon the quick and facile inhibition of old automatic habits.

4. Even though no relationship appeared to exist between social readjustment and successful mirror drawing, it might still be worth while to observe the examinee in a situation where he is *unexpectedly* confronted with a difficult task.

The mirror drawing was carried out in the manner described in Whipple's *Manual of Mental and Physical Tests*. A six-pointed star, shielded from direct view, was traced by the examinee while watching his performance in a mirror.

In evaluating the performance it has not been necessary to use methods of great precision. Those for whom the drawing was difficult were slow. Their lines were very heavy, very light, or alternately very light and very heavy. Their digressions from the correct path were numerous and wide. They wanted to stop trying. They found themselves unable to move their

hand, or, at best, able only to move it through slight and aimless excursions. Diffuse energy was expended in many pointless ways. They laughed, sighed, squeezed their pencils more tightly, and thought out alibis. Those for whom the drawing was easy worked quickly, uniformly, without waste energy, persistently, and without alibis. In order to avoid equivocal measurements, we have simply arranged the performances in a scale of relative success, in all series discarding the doubtful middle group. In no instance has anyone re-sorting our cases placed a "good" performance in our "poor" group, or *vice versa*.

Figures 1, 2, and 3 are reproductions of mirror drawings which are representative of the most competent, the mediocre, and the least competent performances. These particular examples were selected by Donald T. Griffin and the writer as being characteristic of the first, fifth and sixth, and the tenth deciles of 200 unselected performances in the New Jersey reformatory at Annandale.

The New Jersey Reformatory at Annandale (for young men) is an open institution, for the better type of offenders, needing and having few facilities for handling a recalcitrant inmate. Those who are committed to Annandale and fail to conform to the disciplinary requirements there are transferred to other institutions more suitable for their custody and discipline. Two hundred consecutive cases were examined and an unquestionably "good" and an equally "poor" group were selected from these (40 in each group) on the basis of their mirror drawings alone. In contrast to the one boy transferred from Annandale who belonged to the "good" group, there were five transferred who belonged to the "poor" group. This difference becomes particularly impressive if one is aware of the large amount of trouble an inmate must make before his transfer is considered. Comparing the "poor" with the "good" group, we find that the former had half again as many reports recorded against them for infractions of discipline and showed the

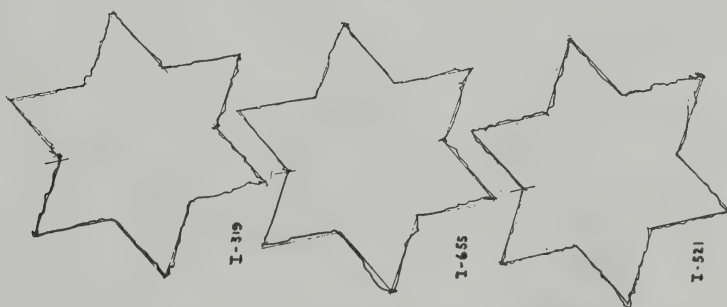


FIGURE 1

MIRROR DRAWINGS TYPICAL OF THE MOST COMPETENT PERFORMANCE

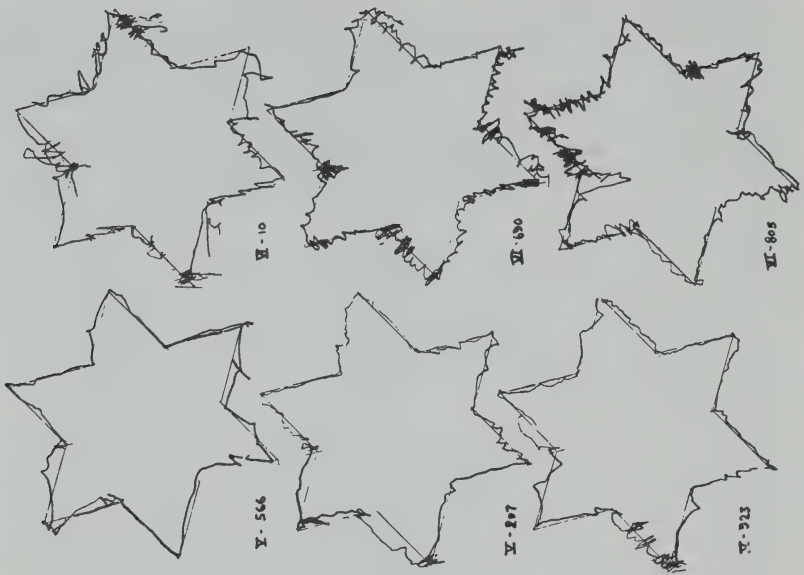


FIGURE 2

MIRROR DRAWINGS TYPICAL OF THE MEDIOCRE PERFORMANCE

same differential with respect to the number of their arrests prior to commitment to the reformatory.

At the State Home for Girls (for delinquents under 16 years of age) a number of parole violators were studied. Only those were selected who were well known to the psychiatrist, school teacher, psychologist, and parole

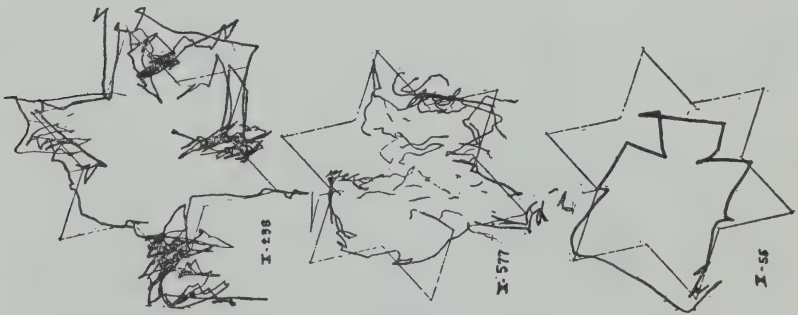


FIGURE 3

MIRROR DRAWINGS TYPICAL OF THE LEAST COMPETENT PERFORMANCE

TABLE 1

	Mirror drawing	
	Good	Poor
Number of cases transferred from each group to the State Hospital at Rahway	0 1	3 2
Total number of arrests prior to commitment (for each group of 40 cases)	74	104
Total number of institutional infractions of rules (for each group of 40 cases)	37	52

supervisor. On the basis of mirror drawings alone, 12 were selected who were unquestionably sufficiently "poor" and 12 who were sufficiently "good" that no one in rearranging the groups would confuse any of the records. Each official mentioned above, of course without knowledge of the mirror drawing, divided the list so that there would be two equal groups. The first contained the half who, considering all the factors involved, might be expected under any circumstances of parole to be more liable to return because of inherent inability to avoid making their old mistakes. The second group contained those whose parole violation might more properly be ascribed to an unfortunate environmental accident, and who might under favorable conditions have been expected to stay out of the institution. Two out of 12 of those whose mirror drawings were "good" were found to be in the group who would not have been given a good parole prognosis. Ten of those whose mirror drawings were "good" were given the better parole prognosis. Although a coincidence of this type might happen by chance, according to a liberal calculation, about one in 27 such series, it seems highly doubtful that this is the twenty-seventh time.

These data are not presented to controvert any psychological principle of long and reputable standing. On the contrary, they are distinctly in line with old and unequivocal findings. The use of the method dates at least as far back as the last decade of the last century and probably farther. Whipple's *Manual* quotes Dr. Weidensall as follows:

"This test isolates better than any we have tried at Bedford those who are incapable of sustained effort under difficulties. . . . Chiefly, however, is the test of interest in the case of those who are bright enough to trace the star well, but too unstable to do so. These are invariably the girls who are difficult to manage in the institution."

Quoting Whipple further:

"The first of these tables shows that, both in the first and last trial, and whether maximal, minimal, median, average, or upper or lower quartile is considered, the three groups are in-

variably arranged in the same order—students best, Bedford women last and the maids intermediate.”

Furthermore, even some of the “discrepancies” reported in the literature might not appear so serious had mirror drawing been considered as an examination method, rather than as an intelligence test. Burt, who reports a correlation of .67 between speed in mirror drawing and intelligence as estimated by teachers, is perhaps not proved to be in error by Calfee who finds no correlation between mirror drawing and the scores made on intelligence tests. For, after all, the teacher does estimate highly the intelligence of the child who quickly and easily inhibits those bits of behavior not approved by the teacher.

The results presented here in relation to the history of the method and in relation to current practices are surprising only if one disregards or underestimates the efficacy of propaganda in determining psychological usages. Never more universally or vociferously than now has the psychologist been assured, by influential non-psychologists, that his function is the determination of the mental age, and that alone. Only by a rigorous application to social problems of all the useful methods developed in experimental psychology can the psychologist hope to maintain the respect of his colleagues in the social sciences. The mirror drawing herein described is only one of the many possibilities.

These results are of further value in that they show the utility of an examination method when used to record photographically the mental picture at a given moment, whether or not the method yields a genuine measurement, or conforms to certain canons by which one rates the value of a mental test. Mirror drawing is not “reliable” either by the split-half technique or by retest criteria. Neither are many useful photographs. Its use does not necessarily imply a distinction between capacity and performance, nor does it imply structuralistic or functionalistic concepts of mind.

A skilled clinical psychologist who does not expect more information from the use of mirror drawing than any examination method is *certain* to yield can obtain a mental photograph of a delinquent examinee which, studied in connection with other factors, may lead to a considerably earlier identification of a criminal recidivist than would be possible without the use of mirror drawing.

*New Jersey State Department of Institutions and Agencies
Trenton, New Jersey*

SOME STATISTICS DERIVED FROM RECENT QUESTIONNAIRE STUDIES RELATIVE TO HUMAN SEXUAL BEHAVIOR

O. L. HARVEY

The table of statistics presented with this note constitutes an attempt to summarize some of the data reported in recent questionnaire investigations dealing in part with the incidence of various forms of sexual behavior among men and women of relatively superior intellectual and social status.

Initially, the problem was to reduce the various findings to a comparable basis. Suppose, as an illustration of the problem, that, of 1000 persons circularized, only 200 respond, of whom 20 admit certain practices, 5 of them commencing before 14 years of age. It will depend upon the investigator's universe of discourse whether that five be reported as $\frac{1}{2}\%$, $2\frac{1}{2}\%$, or 25% of the total. But not all investigators assume the same universe of discourse. Consequently, the comparison of percentages usually quoted sometimes leads to confusion.

The second problem was to determine the actual number of persons proper to any given category. Not all investigators provided precise data for every separate age level; some stated the frequency for a class-interval covering three or four consecutive age-levels. To adapt such types of distribution to some other which is deemed more satisfactory as a basis of comparison common to all of the related studies involves interpolation of essentially a rough nature. Again, in two studies, not only are percentages alone quoted, but also only a vague indication is given concerning the extent of the total population from which those percentages were derived (e.g., "some three-quarters of 250"). Readjustment of such vague data to conform to a basis of division common to a group of similar studies obviously must involve guesswork.

Thirdly, the fact, that the definitions of terms used vary from one investigator to another, introduces an uncontrollable complicating factor into the situation. By way of illustration, although the percentages of those admitting masturbation, reported in two of the most carefully defined studies, appear to be closely similar, actually the definitions used are noticeably different. The one excludes "infantile" forms of masturbation by a process of positive definition; the other does so by implication only.

Finally, the nature of the sampling will affect the percentages reported. To illustrate by way of extremes, whereas one study relates to boys of average age 16, another deals with married men averaging about 42 years of age. Again, whereas one sample population exceeds 1000 women, another amounts to only 47. Casual comparison of percentages, thus subject to qualification, are obviously liable to lead to disagreement and misunderstanding.

In view of the fact that it is on the basis of studies such as these that popular statements concerning the incidence of certain practices are based,

TABLE 1
INCIDENCE OF VARIOUS FORMS OF HUMAN SEXUAL BEHAVIOR*

Investigation <i>a</i>	Date <i>b</i>	Sub- jects <i>c</i>	Age <i>d</i>	Issued <i>e</i>	Completed <i>f</i>	Admissions				Cumulative age-frequencies: Auto-(masturbation)				Cumulative age-frequencies: Heterosexual				Reference pages		
						Auto	Hetero	Homo	Homo	—12	—15	—18	—21	—24	—12	—15	—18		—21	—24
1. Hamilton (married)	1929	M	42	100	100	88 (88)	59 (59)	—	—	21 (21)	57 (57)	73 (73)	81 (81)	82 (82)	—	—	—	—	350-1, 427-8, 492, 508	
2. Peck and Wells	1923	M	23	188	188	139 (74)	66 (35)	—	—	15 (8)	75 (40)	123 (67)	133 (72)	—	—	2 (1)	6 (3)	21 (11)	45 (24)	697, 702, 707
3. Peck and Wells	1925	M	23	238	238	195 (82)	88 (37)	—	—	43 (18)	138 (58)	179 (76)	190 (80)	194 (82)	2 (1)	5 (2)	30 (13)	67 (28)	83 (35)	502, 514, 517
4. Achilles Group B	1923	M	22	108	42	27 (64)	20 (48)	—	—	6 (14)	19 (45)	26 (62)	27 (64)	—	2 (5)	4 (10)	6 (14)	11 (26)	16 (38)	8, 50
5. Exner	1915	M	19	562	531	319 (60)	186 (35)	—	—	11 (21)	306 (58)	431 (81)	—	—	—	161 (30)	170 (32)	—	—	4, 16-18
6. Achilles Group D	1923	M	16	544	406	190 (47)	23 (6)	—	—	10 (2)	110 (27)	189 (46)	190 (47)	—	3 (1)	7 (2)	21 (5)	23 (6)	—	8, 50
7. Hamilton (married)	1929	W	42	100	100	40 (40)	35 (35)	27 (27)	—	13 (13)	17 (17)	23 (23)	31 (31)	34 (34)	—	—	—	—	—	427-8, 493, 508; Davis, 304
8. Davis (married)	1929	W	38	3000	1000	381 (38)	71 (7)	157 (16)	—	153 (15)	219 (22)	252 (25)	282 (28)	305 (31)	—	—	—	—	—	xii, xiii, 102-3, 212, 232
9. Davis	1929	W	37	3000	1000	603 (60)	105 (11)	184 (18)	—	261 (26)	322 (32)	345 (35)	379 (38)	405 (41)	—	—	—	—	—	ix, xviii, 80, 155-6, 232-3, 298, 304
10. Achilles Group B	1923	W	21	97	47	13 (28)	7 (15)	—	—	6 (13)	8 (17)	10 (21)	11 (23)	13 (28)	2 (4)	2 (4)	2 (4)	4 (9)	5 (11)	2, 8, 50
Approximate median of percentages						(70) (40)	(35) (20)	—	—	(15) (15)	(50) (20)	(71) (25)	(75) (30)	—	(3)	(5)	(12)	(20)	(33)	

*All percentages have been determined relative to the totals given in Column 1. Percentages are indicated by means of parentheses.

EXPLANATION OF TABLE

Column *a*—name of investigator. Column *b*—date of publication of investigation. Column *c*—sex of subjects replying to questionnaire. Column *d*—average age of subjects replying to questionnaire. Column *e*—Number of questionnaires issued. Column 1—number of questionnaires returned completed. Column 2—number and percentage of subjects clearly admitting previous autosexual practices (masturbation). Column 3—number and percentage of subjects clearly admitting previous heterosexual intercourse (prior to marriage). Column 4—number and percentage of subjects clearly admitting previous homosexual practices of a frankly sexual nature. Columns 5 to 9—numbers and percentages of subjects clearly admitting masturbation commenced prior to the age indicated at head of column. Columns 10 to 14—numbers and percentages of subjects clearly admitting heterosexual intercourse commenced prior to the age indicated at head of column. Column 15—pages in original publication for which this information was derived.

and because unqualified generalizations are often made, irrespective of the source of the original data, it is desirable that the statistics quoted should at least represent as nearly as possible the consensus of findings from available studies. Consequently, in the appended table, percentages representing the medians of the studies concerned have been determined, and are stated in the last two rows of the table. They have been expressed only approximately so as to facilitate recall and also to avoid any unwarrantable appearance of preciseness.¹

The following notes, particularly relevant to the table appended, are submitted at this stage:

a. Numbers in the body of this table enclosed in parentheses represent percentages; all others represent actual frequencies.

b. Column 3 includes "sexual intercourse with future spouse"; Column 4 excludes cases with only mildly homosexual tendencies.

c. The percentages indicated represent only those cases clearly and definitely admitting the practices concerned. Replies which, for one reason or another, were unclassifiable have been omitted. The percentages reported, therefore, represent minima.

d. The percentages were calculated, in every case, on the basis of the population responding to the questionnaire as a whole. They have not been calculated (a) on the basis of the total population canvassed, nor (b) on the basis of the population responding to the individual item alone.

e. The data as here tabulated are often only approximate. To reduce all studies to a comparable basis, it was necessary to engage in a certain amount of shrewd guesswork, some interpolation, and smoothing of curves. It would be much appreciated if the authors of the original publications here referred to would publicly correct any gross errors which appear in these approximations.

REFERENCES*

*References are numbered to correspond to the order given in Table 1.

1. HAMILTON, G. V. A research in marriage. New York: Boni, 1929. Pp. xiv+570. (Married men.)
2. PECK, M. W., & WELLS, F. L. On the psychosexuality of college graduate men. *Ment. Hygiene*, 1923, 7, 697-714. (Single men.)
3. ————. Further studies in the psychosexuality of college graduate men. *Ment. Hygiene*, 1925, 9, 502-520. (Single men.)
4. ACHILLES, P. S. The effectiveness of certain social hygiene literature. New York: Amer. Soc. Hygiene Asso., 1923. Pp. 116. (Group B, single men, University extension class students.)
5. EXNER, M. J. Problems and principles of sex education. New York: Asso. Press, 1915. (Single men; excludes those cases not questioned by the investigator.)

¹To appreciate the difficulties involved in resolving studies of this nature, it is interesting to compare the findings reported here with those reported by Davis (8, p. 105).

6. ACHILLES, P. S. See (4). (Group D, single men, high-school boys.)
7. HAMILTON, G. V. See (1). (Married women.)
8. DAVIS, K. B. Factors in the sex life of twenty-two hundred women. New York: Harpers, 1929. Pp. xxii+430. (Married women.)
9. ————. *Ibid.* (Unmarried women.)
10. ACHILLES, P. S. See (4). (Group B, single women, University extension class students.)

Austin, Texas

PLEASING PERSONALITY

EDWIN G. FLEMMING

When it is said that a person has personality, it generally is meant that he has a pleasing personality. Now, what is a pleasing personality?

In reading the psychological literature one gets the impression that there is some confusion in the concepts of personality among the psychologists. Some seem to think that when they are investigating character that they are engaged in research in the field of personality. Others seem to have the impression that social adjustments are related to personality. Again, tests of emotional balance are supposed to show something about the personality, while in other cases emotional expressiveness or inexpressiveness is linked with concepts of personality, since, according to the psychoanalysts, it is good for the individual to express his emotions.¹

The purpose of this investigation is to determine the relation of pleasingness of personality to emotional steadiness, to emotional expressiveness, and to social adjustment.

The criterion of a pleasing personality was the judgment of the fellow students of the subject. The procedure of getting the judgments departed somewhat from the usual procedure. Since it would have been impossible to get from three to five judges who knew all the subjects well enough to rate them, it was determined to have everybody rate everybody else on a definite and circumscribed scale, and to take the average rating for each subject.

To be more specific, the names of all the students taking the course in elementary psychology—a total of about two hundred—were mimeographed, and opposite each name were the numbers ranging from one to ten. Every student was given such a list with an instruction sheet worded as follows:

Rate the following individuals on how they affect you. Do not consider their general reputations. Try to answer the question: "Is my response to this individual pleasant or unpleasant?" If your response is very pleasant, draw a circle around the figure

¹For a more detailed indication of the confusion existing see the brief survey of the use of three tests in this field given in (1).

ten (10); if your response is unpleasant in the extreme, put a circle around the one (1). In order to gauge the pleasantness of your response, consider the response that you make to the individual you like best as the standard for circling the ten (10). In gauging the unpleasantness of your response, consider the response you make to the individual you dislike the most as the standard for encircling the one (1). Rate the rest with relation to these two extremes. **DO NOT RATE ANY INDIVIDUAL WHOM YOU FEEL YOU DO NOT KNOW WELL ENOUGH TO RATE.**

The number of ratings thus secured ranged from not any for some subjects to over 125. In no case was any subject used who did not have at least five ratings. In all cases the average rating was taken as the estimate of his pleasingness of personality.

The same thing was done with emotional steadiness, in which case the instructions were:

Rate the following individuals on the steadiness or evenness of their emotional responses. Try to answer the question: "Is he erratic or uncertain; is it impossible to tell how he will respond emotionally to any situation?" Consider only actual evidence of his behavior. Consider the most erratic and unsteady individuals you know as rating ten (10); and the steadiest individual you know as rating one (1). Rate all the others with these two as the standards for extremes. **DO NOT RATE ANY INDIVIDUAL WHOM YOU FEEL YOU DO NOT KNOW WELL ENOUGH TO RATE.**

For emotional expressiveness the instructions were:

Rate the following individuals on the basis of their emotional expressiveness. If you consider the individual to be emotionally expressive to a high degree, draw a circle around the ten (10). If you consider the individual to be phlegmatic or emotionally unexpressive to a high degree, draw a circle around the one (1). Take into consideration only actual evidence that you have of the expressiveness or lack of expressiveness; do not consider what you think him capable of. Also do not consider his "acting" ability. As a guide to your estimates, take the most emotionally expressive individual that you know as representing ten (10), and the most inexpressive individual as represented by one (1). Rate the rest accordingly. **DO NOT RATE ANY INDIVIDUAL WHOM YOU DO NOT KNOW WELL ENOUGH TO RATE.**

And with adjustment the directions were as follows:

Rate the following individuals on the basis of whether you

consider them well adjusted or poorly adjusted to the usual situations of life. "Well adjusted" means that he "fits in well" into the various situations of life; in social situations as well as in academic environment; with old people as well as with people of his own age; with individuals of the opposite sex as well as his own sex; with people of importance as well as with those of not much importance; and so on. "Poorly adjusted" means that he does not "fit in well"; that he is "queer"; that he appears to be uncomfortable in all or some of these situations. Base your judgments on actual evidence only; do NOT estimate what you think the individual might do. Rate the most adequately adjusted individual as ten (10); and the most poorly adjusted individual as one (1). Place the others along the scale in the positions where you think they belong. DO NOT RATE ANY INDIVIDUAL WHOM YOU FEEL YOU DO NOT KNOW WELL ENOUGH TO RATE.

The subjects, as indicated, were young men and young women who were taking the elementary course in psychology. They were mostly sophomores, with a sprinkling of juniors and seniors. There were no freshmen. The rating sheets were distributed in December before the Christmas recess. The subjects, consequently, had an opportunity to know each other for over a year, and in no case for less than three months. Since they were cautioned in all rating sheets not to rate anyone whom they did not know well enough to rate, the probabilities are that no ratings were given except with a certain measure of confidence in the ability to make the rating. An examination of the rating sheets also indicated that the judges acted with discretion. In some cases some of the students handed in rating sheets without rating anybody on the specific trait. This happened in all the ratings except pleasingness of personality, on which some individuals were rated by everyone, however few.

In practically all cases the ratings of individuals showed a wide spread, ranging often from one to ten. The averages, however, may be considered really discriminatory, since there were so many judgments in the majority of cases. The average number of estimates given on pleasingness of personality was 51.3; the range was from 7 to 134; and the sigma was 24.9. With steadiness of emotional response the average number of ratings for each subject was 34.9; the range was from 5 to 102; and the sigma was 18.4. The average number of ratings on expressiveness was 34.1; the range was from 3 to 102; and the sigma was 18.3. While on the estimates of adjustment the average number of ratings was 37.0; the range was from 5 to 107; and the sigma was 20.4.

The judgments of the men and women seem to coincide quite well; in no case is there a reliable difference between the average ratings given by the men and the average ratings given by the women. Table 2, for example,

TABLE 1

SHOWING THE AVERAGE NUMBER OF RATINGS ON EACH OF THE TRAITS STUDIED, TOGETHER WITH THE RANGE AND SIGMA

Trait	Average number of ratings	Range of number of ratings	Sigma
Pleasingness	51.3	7-134	24.9
Steadiness	34.9	5-102	18.4
Expressiveness	34.1	3-102	18.3
Adjustment	37.0	5-107	20.4

shows the differences between the average ratings by the men and by the women. While ratings were on a scale of ten, they were all multiplied by ten before being used.

If one considers those ratings on which the difference was at all greater than the sigma of the difference, it appears that men consider men more pleasing than women do; and women consider women less steady in their emotional responses than men do. In neither case, however, is the difference statistically reliable.

Although the men and women agree quite well on the average in their ratings of each other, there appear to be some differences between men and women on the several traits which are worth considering. Table 3 shows those differences.

Table 3 reveals some interesting differences, the clearest of which is that women are, on the average, more expressive than the men; and both the men and the women agree on this difference. According to the judgments of the men, 77% of the women reach or exceed the median for men in expressiveness; and, according to the estimates of the women, 76% of the

TABLE 2

SHOWING THE DIFFERENCES IN THE AVERAGE RATINGS MADE BY THE MEN AND THOSE MADE BY THE WOMEN

Trait	No. of cases	Av. of men's estimates	Av. of women's estimates	Diff.	Sigma of diff.	Diff. divided by sigma of diff.
<i>Ratings on men</i>						
Pleasingness	74	60.59	58.59	2.00	1.76	1.14
Steadiness	61	50.93	50.14	.79	1.58	.50
Expressiveness	57	53.35	54.11	.76	1.62	.47
Adjustment	61	60.30	59.90	.40	2.31	.17
<i>Ratings on women</i>						
Pleasingness	85	63.74	62.56	1.18	1.58	.75
Steadiness	62	49.06	51.53	2.47	1.69	1.46
Expressiveness	61	62.29	61.25	1.04	1.77	.59
Adjustment	62	66.32	64.32	2.00	2.05	.98

TABLE 3
SHOWING THE DIFFERENCES BETWEEN MEN AND WOMEN ON THE TRAITS
INDICATED

Trait	Average on men	Average on Women	Diff.	Sigma of diff.	Diff. divided by sigma of difference
<i>Ratings by men</i>					
Pleasingness	60.59	63.74	3.25	1.68	1.93
Steadiness	50.93	49.06	1.87	1.46	1.28
Expressiveness	53.35	62.29	8.94	1.58	5.64
Adjustment	60.30	66.32	6.02	2.11	2.85
<i>Ratings by women</i>					
Pleasingness	58.59	62.56	3.97	1.66	2.39
Steadiness	50.14	51.53	1.39	1.80	.77
Expressiveness	54.11	61.25	7.14	1.80	3.97
Adjustment	59.90	64.32	4.42	2.25	1.96

women reach or exceed the median for the men. While this represents some overlapping, it is clear that the women are definitely more expressive than the men; and that is true in the judgment of both men and women.

While none of the other differences are statistically reliable, two of them are near enough statistical reliability to be singled out and mentioned. The men consider the women better adjusted than the men. This probably is more or less generally true with college men and women of sophomore rank, since all the psychological evidence indicates that women in later adolescence are somewhat more mature than men of corresponding ages.

The other difference is shown in that women consider women somewhat more pleasing than they do men. This is more or less contrary to the general popular belief, and may probably be accounted for in the difference in the general social status and calibre of the men and women attending the specific college at which the subjects were secured. However, it would be interesting if in other studies it also appeared that women in later adolescence regarded women more pleasing than they did men. Perhaps it reflects some of the unhappy experiences that adolescence has had with irresponsible youth.

Although the differences are not statistically reliable, it is interesting to note that men consider men less steady than they do women, while women consider women somewhat less steady than they do men. This is undoubtedly a reflection of greater familiarity with one's own sex. Perhaps men are under less control when they are with men, but are on good behavior when with women; and, likewise, the women in the dormitories are more familiar with the women in many different circumstances and different moods than the men have the opportunity for. However, the differences probably are of not much significance since in neither case is the difference statistically reliable.

It is also of interest to note that both men and women consider the women more pleasing, more expressive, and better adjusted, although only in the case of expressiveness is the difference really reliable.

While the judgments of men and women agree rather well, on the average, the agreement is not so close in individual cases, as is shown by the correlation technique. In only two cases is the correlation between men's estimates and women's estimates high enough for even a fair degree of accuracy in prediction. Table 4 gives the correlations between the men's estimates and the women's estimates of men and women on the traits under consideration.

TABLE 4

SHOWING THE CORRELATIONS BETWEEN THE ESTIMATES OF THE MEN AND OF THE WOMEN ON THE TRAITS INDICATED

Trait	r between ratings of men and women on men	r between ratings of men and women on women
Pleasingness	.70	.50
Steadiness	.49	.47
Expressiveness	.38	.40
Adjustment	.71	.42

From these correlations it would seem that the women are better judges of men than the men are of women; but there is a fallacy in such an observation, since it implies that we consider the judgment of men of themselves and of women of themselves as the criterion, which is an assumption not necessarily sound. The lack of correlation between the ratings of the men and of the women on individual subjects simply indicates that, for an adequate rating, judgments of both men and women should be secured. Only on pleasingness and on adjustment do both sexes agree to any considerable extent in their judgments of men, while with women you cannot predict with any very high degree of certainty what the other sex will think of them from just the estimates of one of the sexes.

TABLE 5

SHOWING THE INTERCORRELATIONS AMONG THE CHARACTERISTICS INDICATED
Men only.* ($N=89$)

	Steadiness	Expressiveness	Adjustment
Pleasingness	— .46	.27	.70
Steadiness		.25	.08
Expressiveness			.60

*The scores used in all the correlations, both in this table and in Table 6, were based on an average of at least five judgments.

And now we come to the heart of our problem: How is a pleasing personality related to the other characteristics? Table 5, a table of intercorrelations, is the simplest way of showing the relationships that exist.

These correlations indicate that the best measure of pleasingness of personality, other than judgments of pleasingness themselves, is the social adjustment of the individual. Consequently, those who emphasize social adjustments as the major factor in personality probably are on the right track. Expressiveness seems to be of little significance; we apparently like the quiet individual quite as much as the enthusiastic, effervescent person. Steadiness, which probably implies a certain amount of reliability, is also of some value, but not nearly as important as a general social adaptability.

The correlations with women used as the subjects shows a similar situation and corroborates the results with the men. With the women, however, steadiness seems to have a little more weight than with the men. Table 6 gives the correlations.

TABLE 6
SHOWING THE INTERCORRELATIONS AMONG THE CHARACTERISTICS INDICATED
Women only. ($N=118$)

	Steadiness	Expressiveness	Adjustment
Pleasingness	— .52	.12	.67
Steadiness		.43	— .16
Expressiveness			.55

While pleasingness and expressiveness have little relation to each other either with men or women, expressiveness is related to adjustment—the more expressive tending decidedly to be the better adjusted. And, since adjustment is related to a pleasing personality, it would seem logical that anything that contributes to adjustment would contribute to a pleasing personality. But the multiple correlation for men between pleasingness and adjustment with expressiveness added increases the correlation to only .72, which is not a very significant change. Steadiness is of far more importance, since when it is added to the correlation between pleasingness and adjustment the multiple correlation becomes .87.

With women the multiple correlation between pleasingness and adjustment combined with expressiveness is .71; while, when steadiness is added instead of expressiveness, the multiple correlation becomes .79.

In conclusion, it may then be stated that women are positively more expressive than men as determined by the judgments of both men and women; but that expressiveness has little significant relation to pleasingness of personality. On the other hand a pleasing personality means one that is socially well adjusted, and in which, whether expressive or inexpressive, the general tendency of the emotional response can be depended upon.

REFERENCE

1. FLEMMING, E. G. The predictive value of certain tests of emotional stability as applied to college freshmen. *Arch. Psychol.*, 1928, No. 96. Pp. 61.

29 Claremont Avenue
New York City

THE MEASUREMENT OF INCENTIVES AND THEIR EFFECT: A
CONTRIBUTION TO METHODOLOGY AND ORIENTATION
RESULTING FROM THE EXPERIMENTAL USE OF INCENTIVES

CLARENCE LEUBA

Before anything can be measured satisfactorily it needs to be defined, isolated, and controlled; a specific technique of measurement must be worked out.

DEFINITION OF AN INCENTIVE

An incentive is a form of motivation; the latter is the more inclusive term. For practical purposes, we suggest that the use of the term incentive be limited to those motivating situations which can be used fairly universally and frequently in connection with any activity in progress as spurs to more intense and persistent effort. Objectively considered, incentives are situations which can usually be introduced to facilitate any activity in progress. Words or other expressions of praise, disapproval, or recognition are situations which can be used in this manner. A rivalry situation may be present whenever it is possible to have several persons perform the same type of activities and to compare their results. Not only may the individual be stimulated by the imagined prospect of outdoing some or all his fellows, but the perception of others at the same task will usually itself be stimulating.

Positive incentives are situations which an individual will exert himself to bring about or maintain; once achieved, such situations are usually exhilarating and stimulate the individual to further exertion. Thus praise is an incentive in that a person will exert himself to bring about or maintain a situation in which he is receiving praise, and in that the words and expressions of praise themselves stimulate him to further effort. Some incentives are of a negative nature; they are situations, such as words or expressions of disapproval or blame, which stimulate the individual to exertions usually persisting until situations of that sort have been avoided.

Our general conception of an incentive should include not only the objective external situations we have just been considering, but also the attitude or behavior pattern aroused by these situations. This attitude is one of effort or striving; the vitalist would explain it as one of the fundamental

characteristics of mind; the mechanist might consider it as consisting essentially of muscular tensions and a stirred up visceral condition. There would seem to be only *one* general type of incentive attitude—an attitude of more or less mild emotional excitement. We think of a variety of incentives, praise, blame, rivalry, etc., because a number of different situations will excite this incentive attitude. It is from those situations that incentives are named. It is correct, therefore, to speak of rivalry situations, of praise situations, etc., and of *the* incentive attitude stimulated by them.

THE ISOLATION AND CONTROL OF AN INCENTIVE

Whenever it is possible to prevent anyone from knowing his particular performance and that of anyone else, it is easy enough to exclude all incentives from the experimental situation. Incentives will not be present at all if no one knows, or at least if no one pays any attention to, the quality or the quantity of each individual's performance. Incentives cannot be present unless someone can, and does, at least roughly, gauge his performance and compares it with that of others, with his past performances, or with some theoretical standard. Even though many persons of about equal ability may be at work on the same tasks, there can be no rivalry unless there is some knowledge on the part of each of his output and of that of others; nor can there be any praise or blame if no one knows what each has done. Likewise, for group incentives to be present, there must be some knowledge of the output of the group.

To isolate and control specific incentives is a much more difficult, if not impossible, task. Take, for instance, a goal as a specific incentive. Simply having an objective to be reached in a certain minimum of time arouses the incentive attitude of striving. But usually the possibilities of praise, of recognition, and the avoidance of blame accompany the achievement of an objective; though the experimenter may have specifically mentioned only a goal, other incentive situations will be actually present, or imagined, the moment subjects realize that their individual outputs will be known. How shall one determine to what incentives to attribute the increased performance? One subject may be motivated chiefly by the prospect of praise, another by rivalry, another by the objective as such, and still another by a combination of these.

Though a material reward, such as a piece of candy, is *per se* an incentive for most children, it is usually much more than just a pleasant taste to be secured; it may also serve as an objective, as a means of securing recognition and praise, and of outdoing one's fellows, and conversely, of avoiding shame, ridicule, obscurity, or being outdone by others. Again, there is no way of knowing how much of the increased output to attribute to each of these possible incentives. Incentives, other than the material reward, as such, have not been controlled.

It would seem that incentives are interrelated; that, if the experimenter

introduces one into a work situation in which previously there were none, he tends to introduce others also; that, if a subject achieves one incentive, he probably achieves others, too; if he outdoes others, he is probably also in line for some recognition by his fellows and some praise from them. Perhaps the experimenter has consciously promoted only rivalry, but for many subjects the outstanding incentive may have been not the presence of others at the same task or the imagined possibility of outdoing them, but the possibility of securing the approval of the experimenter or of improving their own records, irrespective of others.

Though many experimenters have laid claim to the measurement of the effects of certain specific incentives such as praise or rivalry, actually their results were the outcome of a complex situation, of which the incentive they stressed was but a part. Incentives were really controlled neither in their experimental nor in their control groups.

It might be possible to isolate the effect of some incentive, such as a material reward, by having two groups of subjects, one in which there would be an objective, of no value in itself, as an incentive and one in which there would be introduced both the objective and a material reward. If there were a significant difference between the outputs of the two groups, it might then be attributed to the material reward as such.

THE TECHNIQUE OF MEASURING THE EFFECT OF INCENTIVES

It is to be expected that people will not exert themselves any more than necessary to attain an incentive. In the case of a rivalry situation, people will not wittingly make more of an effort than is required to outdo their rivals; in the case of a praise situation, they may exert themselves only moderately if that suffices to secure praise from a certain person, though, were it necessary, they might be willing to exert themselves much more to obtain his approval.

To give praise indiscriminately to a group of subjects, or even to give praise only to the outstanding members of a group, is not an accurate method of measuring the *full* stimulating power of praise as an incentive. At most, the experimenter has measured the extent to which the subjects think they have to, and care to, exert themselves to get praise from the particular experimenter in question. Had he had higher standards, it might have appeared that they were willing to exert themselves much more to receive his praise.

Within limits, which vary probably with the individual, the more difficult the achievement of an incentive is made, the more will individuals exert themselves to attain it; and, conversely, the easier its attainment is made, the less will be its stimulating power. If an incentive such as praise is achieved, no matter what the performance may be, it ceases eventually to have any exciting effect at all.

Though incentives are to be carefully distinguished from organic drives,

these two types of motivation are alike in that the more easily available are the means of satisfying an organic drive, such as hunger, the less the rôle of that drive as an organic motivating factor in the life of the individual.

Dr. F. H. Allport has suggested a technique for measuring the full stimulating power of an incentive; it is a method similar, in some respects, to that used in psychophysics for the measurement of sensory thresholds. The requirements for achieving an incentive, such as social recognition, are increased step by step from a minimum, at which everyone receives the incentive, to a maximum requiring an effort beyond that which anyone is willing to make to obtain the incentive; then the requirements are gradually decreased from the maximum at which no one received the incentive to the minimum again at which everyone received it. The average of the maximum performance of an individual, during the increasing requirements, and of his maximum, during the decreasing requirements, is then a measure of the extent to which he would exert himself rather than do without the incentive in question. It is a measure of the full stimulating power of the incentive.

It is desirable to use, as experimental material, some task in which practice effects have come to an end; the subjects should have already, under suitable motivation, repeatedly reached their physiological limit. Most experimenters have been interested in the accelerating effects of incentives upon the learning process, though those are not the conditions under which incentives can most readily be controlled. Such conditions render it difficult to disentangle practice from incentive effects. If an old skill be used in which the subjects have already reached their physiological limit, any change in output can be attributed to changes made in the work setting by the experimenter.

It is essential that the work be of such a nature that the same method must always be used in its performance.

DESCRIPTION OF AN EXPERIMENT TO MEASURE THE EFFECTS OF A QUANTIFIED INCENTIVE

Since the foregoing analysis of incentives and of a technique for their isolation and measurement was evolved during the course of our own experimentation, first with adults and then with children, none of our experiments conform completely to the methodology we have described. Nevertheless, as our experiments differed in several aspects from previous ones in this field, we will report one of them briefly.¹

The object of this experiment was to compare the average performance of a group of children, when incentives were excluded from their work setting, with that which they would make, if necessary, to obtain a known

¹A complete and detailed account of the experiment will be found in the *J. Abn. & Soc. Psychol.* for Nov., Dec., 1930.

amount of an objective incentive situation, such as one five-cent chocolate bar. The material consisted of simple, two-place multiplication problems of uniform difficulty. The 35 subjects were members of Class 5A at a public school in Syracuse, New York. Their average age was 11.3 years. They were told that, for a while, they would take part three times a week in a ten-minute multiplication practice period in the classroom adjoining their own. The experiment consisted of 21 ten-minute work periods and lasted two months. The first two work periods were of a preliminary nature, that the children might become adapted to the general setting of the experiment; then followed four trials without incentives, it being made clear to the subjects that no one counted the problems they had done or even looked at them. There were trials, without incentives, or, as we shall call them, no-incentive trials, before and after the introduction of an incentive.

Following the first no-incentive trials we ran a set of five trials, before each of which the experimenter announced that he had a box of five-cent chocolate bars and that those who did at least the number of problems marked on their sheet would receive a bar at the end of the ten-minute practice period. The requirements for the chocolate bar, in terms of number of problems to be done within the stated time, were different for each subject, depending upon his past performance; they were increased from one trial to the next. The amount of increase in the requirements was not determined upon at the start of the experiment but was so arranged empirically for each subject from one trial to the next that an equal proportion of the slow, medium, and fast multipliers satisfied them on each trial.

The experimenter was surprised by the increase in the speed of multiplication made by the subjects to attain a chocolate bar as the requirements for it were progressively raised. He made the requirements on the first two trials purposely easy, so that the subjects would have the experience of getting and eating the chocolate, but, had he realized the magnitude of the gains which the subjects were capable of making, and willing to make, to secure the chocolate, he would have increased the difficulty of the requirements more quickly and radically. Theoretically, this series of trials should have been continued with increasingly difficult requirements until no one secured the chocolate. However, for practical reasons, it was stopped after five trials; signs of overnervousness appeared in some; others were becoming restless and dissatisfied as they no longer secured the chocolate while others still got it. It seems certain that some of the 35 subjects had not quite reached the maximum output which they would have made, if necessary, to get the chocolate.

After a set of three trials without incentives, chocolate bars were again used as an incentive. This time the requirements for the chocolate were used in the order of decreasing, instead of increasing difficulty, until nearly everyone obtained a chocolate bar.

After a last set of no-incentive trials, the subjects were given the final

two trials of the experiment using all the incentives the experimenter could think of; rivalry, praise, and recognition being introduced as well as the chocolate.

For each subject we calculated a chocolate performance index equal to the average of his highest output when the requirements for the chocolate were increasing to a maximum, and of his highest output when the requirements for the chocolate were decreasing from that maximum. This index indicated the number of problems he would do, if necessary, to secure a chocolate bar. The average chocolate performance index for the lowest quartile showed a gain of 92% over their level of work when incentives were absent; for the highest quartile the gain was 32%; the average gain for the group, taken as a whole, was 52%. Though the slow multipliers gained the most proportionately (in terms of percentage,) they tended to gain the least absolutely (in terms of number of problems).

There was a surprisingly constant level of performance during all three of the sets of no incentive trials, though trials in which an incentive had been used intervened. The average for each no-incentive set varied by only one-half problem from 23.4 problems. There seemed to be a low and fairly constant no-incentive level of work for the group.

The subjects included 19 boys and 16 girls; the boys averaged slightly fewer problems than the girls during the no-incentive trials but had slightly higher chocolate performance indices. These differences were too small to be reliable. They are in harmony, however, with the common observation that, on the whole, elementary-school work, as such, has less stimulating power for boys than for girls; and with the general principle that those tend to gain the most in output with the introduction of incentives for whom the work as such has the least stimulating value.

To prevent a condition from arising in which the subjects would at times no longer perform the operations involved in multiplication, and, instead, write down random figures, we had to draw attention to errors, especially during the incentive trials, and even inflict penalties for excessive ones (over 30% wrong). Though the experiment was a speed, and not a quality, test it was essential that the subjects always actually perform the multiplication operations.

FUTURE STUDIES OF INCENTIVES

That incentives, in general, usually have an accelerating influence, both on the speed of learning and on the performance of a familiar task, has been frequently demonstrated, both in and out of the psychological laboratory. Alluring as is the goal of more refined and specific quantitative measurements of incentives and of their effects, we had perhaps better dismiss it, for the time being at least, in favor of basic qualitative experimentation. Since the variables are so numerous and so ill-defined and interrelated, quantitative measurement might at present serve chiefly to give an unwarranted

impression of exactness and to becloud the urgency for preparatory, qualitative experimentation. We have seen that, because of the extent to which incentives are interrelated, it is well-nigh impossible to isolate one for measurement and to control the others; that results depend in some degree upon the personality of the experimenter administering the incentive; and that, though it may be possible to quantify an incentive, such as chocolate, and to say these results are due to one chocolate bar, these to two chocolate bars, such a procedure would be more than difficult with other incentives, such as praise. Furthermore, whatever quantitative results we did get would doubtless be a function of the frequency with which, or the length of time over which, an incentive had been used, and of the past training of the subjects. The stimulating power of an incentive for any individual is not immutable; it is probably subject to change under varying social conditions.

The following possible avenues of investigation suggest themselves as leading to a clearer understanding of the nature of that form of motivation we have called incentives, and as leading eventually perhaps to the possibility of exact measurement:

1. Studies of the incentive attitude aroused in individuals by incentive situations: A possible hypothesis is that this attitude is similar in many respects to that prevailing during the emotion of anger, except for the specific impulse to strike, or during fear except for the specific impulse to flee; in other words, that it is an undifferentiated emotional excitement of varying degrees of intensity. This incentive attitude might profitably be studied by the same physiological methods which Cannon has used with such brilliant results for the understanding of the emotions of fear and anger. There is need for an objective instrumental method for measuring the muscular tensions which undoubtedly form a part of the incentive attitude. This study would consist essentially of an investigation of the bodily condition present during effort or striving, and would be primarily physiological in nature.

2. The development of incentives in childhood: a study of when and how situations such as approval, praise, recognition, and the presence of others at the same task (rivalry) acquire power to stimulate the child to effort and concentration; this would be a study of the genesis of the stimulating power of incentive situations. The nature of incentives and the manner in which they are interrelated leads us to suspect that they may be portions of total situations which were stimulating to the young child inherently and as wholes; various portions of these situations may then have acquired the stimulating value formerly possessed only by the whole; thus the number of incentive situations multiplies as the child grows older.

Such a study should include an investigation of the technique of controlling the relative stimulating power of various incentive situations. It is possible that we today overdevelop the competitive situation as an in-

centive; witness the tree-sitters and the marathon-dancers; one person need only do something unusual for someone else to be excited to outdo him! The control of the relative stimulating value of incentives is basic for the success or failure of present or proposed economic systems.

Antioch College
Yellow Springs, Ohio

SEX DIFFERENCES IN LEARNING TO WALK A TIGHT WIRE^{1 2}

GRANVILLE B. JOHNSON

RELATION OF THIS EXPERIMENT TO A PREVIOUS EXPERIMENT

This is a continuation of a study in learning to walk the tight wire (1). The original experiment had to do with men only. There were two groups of men, one using a high wire (72 inches above the floor), and the other using a low wire (36 inches). The Army Alpha Intelligence Test was given to both groups, and correlations were made between ability to walk a tight wire and intelligence scores of the subjects. To a limited extent, the problem of age in relation to this learning ability was considered, as also was the emotional factor of fear.

The present problem is an attempt to use women in the experiment on the low wire and to check what differences may exist. One hundred thirty-five women entered upon the experiment; 101 of them finished the problem. Thirty-four were still working on it when the experiment was closed at the end of the school year. These women are to be regarded as a representative group of university women, and, as such, they present a true cross-section of active young womanhood. They had absolutely zero knowledge of tight-wire walking, and entered into the experiment because of the novelty and fascination of learning to walk the wire, and because of a desire to help the Department of Physical Education in this piece of research. Before the experiment was completed, the Army Alpha Intelligence Test was given to all of these women.

METHOD OF CONDUCTING THE EXPERIMENT

At the beginning the student was told to take up the balancing pole, mount to the starting platform, and walk across the wire. The problem for the subjects was to successfully cross the wire three times in succession. The difficulty in walking the tight wire would seem to be lack of keen adaptation

¹This study has been materially strengthened by the painstaking assistance of Professor Thomas R. Garth for which the writer makes grateful acknowledgment.

²This paper was presented before Section I, A.A.A.S., Philadelphia, 1928.

necessary to enable one to control his body sufficiently well to walk upon an object as small as the wire, coupled with the phenomenon of learning to absorb the ensuing vibrations imparted to the wire from the body of the subject. Each subject made three attempts a day and the average of these three attempts constituted a "trial." The trials were conducted on Monday, Wednesday, and Friday of each week, so that the time between trials would be approximately equal. After the trial, the subjects were asked to give an account of their mental, and particularly of their emotional, states while performing, as a means of getting some line on the factor of fear.

Every effort was made to standardize conditions at all times. For instance, the apparatus was constantly checked for wire tension, and no controllable distractions were permitted. The spirit of the young women was sustained throughout the entire experiment, and much credit is due them for the success of this experiment.

The data in the present study was handled in the same way as in the earlier study. The learning ability of the subjects is represented in the number of trials necessary to successfully walk the wire—that is, to walk three times in succession, a perfect score, and by the learning curves themselves, expressed in inches walked per "trial" (the average distance walked in three attempts).

DATA FROM THE STUDIES OF MEN SUBJECTS

In the experiment with men, using the low wire, the median number of trials for the group to successfully cross the wire was 38. In the experiment using the high wire, the median number of trials was 50 (1). See Table 1.

A correlation was found to exist between the intelligence score of the subjects and the number of trials in walking the wire of $-.69$. The cor-

TABLE 1

MEDIAN NUMBER OF TRIALS REQUIRED BY MEN AND WOMEN TO SATISFY THE CRITERION OF LEARNING, PERCENTAGE OF OVERLAPPING AND SOME CORRELATIONS

	A Men on high wire	B Men on low wire	C Women on low wire
Median number of trials	50	34	65.5
Overlap on "B"			12.6
Correlation of intelligence score and number of trials	.65	$-.69$	$-.58$
Correlation of intelligence quotient and number of trials	.54	$-.57$	$-.54$
Correlation of chronological age and number of trials	.34	.34	.28

relation between the intelligence quotient and the number of trials was $-.57$. Yet the correlation of the chronological age of the subjects and the number of trials was only $-.34$. This low correlation is due to the great differences of physical development for any one chronological age. The oldest men of the group made a distinct contrast to the youngest men in that their attitude toward the performance involved much more of fear.

The curve of learning was in keeping with most curves of motor learning.

DATA FROM THE EXPERIMENT WITH WOMEN

The median number of trials for the women was 65.5. The correlation between the number of trials and the intelligence scores was $-.58$. The correlation of the intelligence quotient and the number of trials was $-.54$. The correlation of the chronological age and the number of trials was $.28$, but the subjects did not present a wide enough range of ages to validate any conclusion. These data are tabulated in Table 1. The women indicated subjectively a much greater effect of fear; in fact, it was quite generally a dominating inhibition.

The second line of evidence is found in the overlapping of the women's distribution of the number of trials necessary to learn to walk the wire. This is found to be 12.6%. That is to say, the percentage of women attaining and exceeding the median of the men on the low wire is 12.6. This indicates a significant difference between the number of trials required by the women as compared to the men.

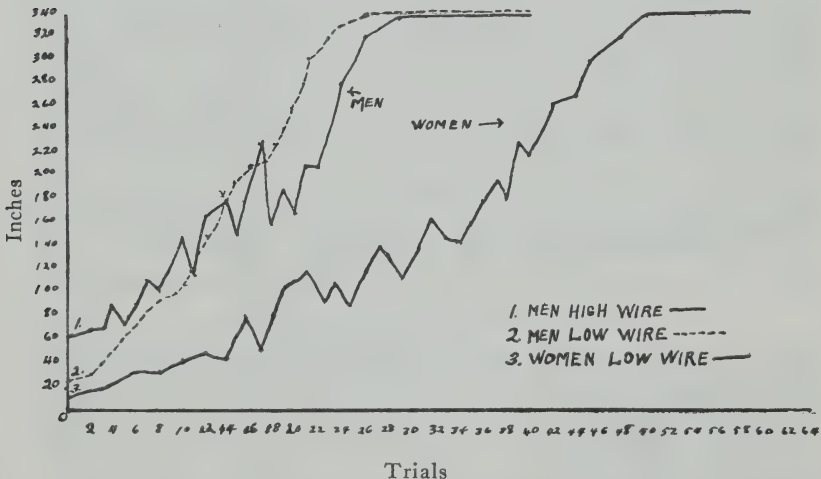


FIGURE 1

LEARNING CURVES FOR MEN AND WOMEN IN LEARNING TO WALK THE TIGHT WIRE

TABLE 2
SHOWING FOR 101 WOMEN THE NUMBER OF INCHES WALKED FOR EACH TRIAL

Trial No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Med. (inches)	10	19	19	25	28	35	37	34	35	42	43	50	55	45	60
<i>Q</i> (inches)	3	12	6	12	14	18	19	16	18	21	21	27	30	23	37
Av. of 5 trials					20.5					36.6					50.6
Trial No.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Med. (inches)	81	60	52	70	83	89	118	120	120	90	110	100	116	120	135
<i>Q</i> (inches)	50	33	35	37	55	60	65	63	70	56	80	71	85	90	91
Av. of 5 trials					69.2					107.4					116.2
Trial No.	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Med. (inches)	150	162	145	150	160	180	185	198	180	225	220	280	283	290	300
<i>Q</i> (inches)	98	99	112	130	125	130	135	139	125	136	143	151	150	155	145
Av. of 5 trials					153.4					193.6					274.6
Trial No.	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Med. (inches)	325	331	340	353	360	360	360	360	360	360	360	360	360	360	360
<i>Q</i> (inches)	141	135	131	140	145	120	121	113	97	90	85	81	50	61	53
Av. of 5 trials					341.9					360					360
Trial No.	61	62	63	64	65	66	67	68	69	70	71	72	73		
Med. (inches)	360	360	360	360	360	360	360	360	360	360	360	360	360	360	
<i>Q</i> (inches)	48	38	48	39	38	38	29	31	25	10	0	0	0	0	
Av. of 5 trials					360					360					

SEX DIFFERENCES IN LEARNING TO WALK A TIGHT WIRE

An examination of the curves in Figure 1 will reveal sex differences on a basis of comparison of the curves for the learning on the low wire. The rate of improvement is seen to be much slower for women. The curve for the men reaches its maximum in the twenty-sixth trial in the low wire, and in the twenty-ninth trial in the high wire, whereas the curve for the women reaches its maximum not before the fiftieth trial after a very slow rate of climbing.

CONCLUSIONS AS TO SEX DIFFERENCES

1. Women of superior intelligence, as measured, mastered tight-wire walking slightly more easily than women of inferior intelligence. The situation is better for men in this respect than for women.

2. Our data are not sufficient to come to any conclusion regarding the relation of age to the learning ability of women. There is a slight tendency on the part of the younger men to be somewhat superior in learning ability to older men.

3. Fear with women was a dominating inhibition, whereas it did not play an important part in the experiment as far as the men were concerned, except in comparison of the high-wire group to the low-wire group.

4. The curves for women and men are decidedly different in form and slant.

5. There is a distinct difference between men and women in gross ability to learn to walk the wire as indicated by the median number of trials alone. Only 12.6% of the women equalled or surpassed the median score of the men on the low wire. This overlapping shows a real sex difference.

REFERENCE

1. JOHNSON, G. B. A study in learning to walk the tight wire. *Ped. Sem.*, 1927, **34**, 118-128.

University of Denver
Denver, Colorado

A NOTE UPON THE RELATION OF ACTIVITY TO SEX AND RACE IN YOUNG INFANTS

KARL C. PRATT

The literature of sex and race differences is all too frequently reflective of the bias of the observer because the methods of study are insufficiently objective. Substitution of the machine as observer and recorder is an aid towards impartiality for then the records so obtained may be presented in such fashion that anyone may evaluate them apart from the view or

interpretation of the investigator. The use of the stabilimeter (2) has made possible objective records of the general bodily activity of newborn infants and thus provided data which may be analyzed from the standpoint of sex and race.

To the casual observer newborn infants appear to differ qualitatively in their movements according to their sex and race. Differences are noted also by the trained observer as is shown by Gatewood and Weiss (1). But differences in specific reflexes may appear qualitatively and not manifest themselves in the measures of total or gross activity.

In this report the infants are for the most part the same as those observed in (1), but the polygraph records of stabilimeter oscillations furnish the data, whereas in the former study the qualitative analysis of movements depends upon the protocols of the observers. The ratio $t/T \times 100$ (where t is the time in seconds that the infant is actually moving and T is the total

TABLE 1
ACTIVITY VALUES ACCORDING TO SEX AND RACE

	Male	Female	All		
Black					
No. of infants	16	16	32		
No. of control periods	112	121	233		
Range	0-100	0-100	0-100		
Mean	22.75	26.35	24.65		
P.E.	1.42	1.17	.92		
$S.D.$	22.40	19.10	20.85		
Diff.			1.95		
P.E.					
White					
No. of infants	15	23	38		
No. of control periods	132	195	327		
Range	0-100	0-100	0-100		
Mean	26.25	26.40	26.35		
P.E.	1.20	.96	.74		
$S.D.$	20.45	19.95	20.10		
Diff.			.09		
P.E.					
All					
No. of infants	31	39	70		
No. of control periods	244	316	560		
Range	0-100	0-100	0-100		
Mean	24.65	26.40	25.65		
P.E.	.92	.74	.58		
$S.D.$	21.45	19.60	20.45		
Diff.			1.48		
P.E.					

time of the control period—during which the infant is left unstimulated) has been selected as the measure of activity. These ratios were secured for all positions of the control period with respect to the experimental period proper and for all physiological conditions (wet, dry, etc.) during which the control period was completed before the infant was returned to the nursery. The age range is from birth to about two weeks.

Table 1 shows:

1. That there is no statistically significant difference in the general activity of male and female infants ($\text{Diff.}_{M's} / \text{P.E.}_{diff.} = 1.48$).

2. That the magnitude of this difference is due to the lesser amount of activity of male negro infants as compared with that of female negro infants ($\text{Diff.}_{M's} / \text{P.E.}_{diff.} = 1.95$), which conforms more closely to that of the whites. The white males and females closely approximate each other ($\text{Diff.}_{M's} / \text{P.E.}_{diff.} = .09$).

3. That there is no statistically significant difference between negro and white infants ($\text{Diff.}_{M's} / \text{P.E.}_{diff.} = 1.44$), the apparent difference being ascribable to the lesser amount of activity among the male negro infants of the distribution.

According to the data presented in this study, we may conclude that sex and race are negligible factors in the general bodily activity of the newborn child.

REFERENCES

1. GATEWOOD, M. C., & WEISS, A. P. Race and sex differences in newborn infants. *J. Genet. Psychol.*, 1930, **38**, 31-49.
2. PRATT, K. C., NELSON, A. K., & SUN, H. K. The behavior of the newborn infant. *Ohio State Univ. Stud. Psychol.*, 1930, No. 10. Pp. xiii+237.

The Ohio State University
Columbus, Ohio

BOOKS

CARL MURCHISON [ED.] *A History of Psychology in Autobiography, Volume I.* (By James Mark Baldwin, Mary Whiton Calkins, Edouard Claparède, Raymond Dodge, Pierre Janet, Joseph Jastrow, F. Kiesow, William McDougall, Carl Emil Seashore, C. Spearman, William Stern, Carl Stumpf, Howard C. Warren, Theodor Ziehen, and H. Zwaardemaker.) Worcester, Mass.: Clark Univ. Press; London: Oxford Univ. Press, 1930. Pp. xvii+516.

CARL MURCHISON [ED.] *Psychologies of 1930.* (By Alfred Adler, Madison Bentley, Edwin G. Boring, G. S. Brett, Harvey Carr, John Dewey, Knight Dunlap, J. C. Flugel, Walter S. Hunter, Pierre Janet, Truman L. Kelley, K. Koffka, Wolfgang Köhler, K. N. Kornilov, William McDougall, John Paul Nafe, I. P. Pavlov, Friedrich Sander, A. L. Schniermann, C. Spearman, Leonard T. Troland, Margaret F. Washburn, Albert P. Weiss, and Robert S. Woodworth.) Worcester, Mass.: Clark Univ. Press; London: Oxford Univ. Press, 1930. Pp. xix+497.

The present confusion of "points of view" in psychology is fully documented in these two volumes by thirty-six different authors. Nineteen of these writers are Americans; there are seventeen foreigners in the list, including six Germans, three Russians, three Englishmen, one Swiss, one Frenchman, one Italian, one Dutchman, and one Austrian. A truly representative group!

And yet, after reading these two volumes, one is at a loss to say what psychology means today as a body of scientific knowledge, or as a field of scientific research. Out of thirty-six writers, only three feel called upon to attempt a reconciliation of the sharp prevailing differences of opinion. Of the three, perhaps the most successful is Boring, who, in his "Psychology for Eclectics," affirms the existence of a real "psyche" to be studied by behavioral and introspective methods, and accepts total structures as "reals" wherever they appear; although the eclectic, he says, "will remain an atomist in his experimentation, because the variables to be correlated in an experiment are essentially discrete" (p. 123).

Nafe and Woodworth also strike the note of reconciliation; but in the main the other writers are intent upon defending their own views against the opinions of others.

As an experimental science, psychology makes, in these volumes, a poor showing. While many of the authors explicitly aim to keep free from philosophical, and especially from epistemological speculation, they do, in truth, present theories to embrace facts, rather than facts that suggest theories. The one notable exception is Koffka with his paper on "Some Problems of Space Perception," for here the *Gestalt*-hypothesis is indicated by de-

monstrable facts. Köhler, too, shows an acquaintance with the laboratory in his revision of a paper that first appeared in *The Psychologies of 1925*; but, excepting the physiologists, it is remarkable that our authors are primarily concerned with a defense of their theories.

The *Psychologies of 1930* consists of twenty-five articles, beginning with one by William McDougall. He defends the "hormic" principle which he finds clearly revealed in the higher mental processes, and by which he proceeds to explain the lower processes, including the inheritance of tendencies to seek goals. The second article, by G. S. Brett, critically expounds "Associationism and 'Act' Psychology." Brett finds in psychological theory a parting of the ways: "One way will lead to a psychology which is scientific but artificial; the other will lead to a psychology which is natural but cannot be scientific, remaining to the end an art" (p. 54).

Harvey Carr's "Functionalism" is a defense of the views commonly associated with the Chicago School. A "function" may be either a "mental activity" or "a service or use for some end." "In experiential terms, an act is a group or pattern of contents exhibiting a unity from the standpoint of its meaningful implications as to end-result" (p. 65); and "while teleological interpretations have been overdone, it does not follow that teleology is essentially non-scientific, and that all teleological interpretations should therefore be discarded" (p. 74).

Miss Washburn, in her "Systems of Motor Psychology," first states her metaphysical background as a dualism of mind and matter, with conscious process as epiphenomena. She proceeds to an interpretation of behavior in motor terms which are mechanistic like those of the behaviorist; still holding, however, in agreement with the configurationists and structuralists, that the description and analysis of mental processes is "legitimate and desirable" (p. 93). One wonders why this concession need be made, since in principle Miss Washburn's motor psychology appears to make no use of epiphenomena.

A "Psychology for Psychologists," by Madison Bentley, is an attempt to deal "with living things as living" under two coordinate groups, the biological and the psychological. The entire being, called the "T-system," consists of a neutral organism (O_n) which indicates the two sciences, "B" and "P," without splitting O_n . The primary category of both these sciences is that of *function*. We have, therefore, the B-functions of metabolism and the P-functions of apprehension. Both are referable to the body, and each has a "coloring," experiential in the case of the P-functions, and physiological in the case of the B-functions.

For Bentley, seeing and hearing are functional facts which are "announced" to the T-system, but "the concept of the environment has no place in psychology," (p. 102). Since "announcement" includes movements as well as things, the distinction of "inner" and "outer" breaks down. "Text" and "context" of the central neural system, rather than "such fixed and ex-

istential couplings as that of organism and environment" (p. 104), set the first sort of determination of function. Distinctions of original and acquired nature are misleading. Every state and movement of a living organism is functionally determined by (a) stock, (b) physicochemical interchange, (c) internal changes (as of new materials), (d) the presence of other T-systems, (e) active association with them, and (f) products of this association. "Upon these primitive functions—not analyzable in terms of sensations, desires, and the like—play the articulations of receptor, brain, muscle, tendon, and gland. Whereupon gradually appear the grosser perception—actions" (p. 108). The author proceeds to list eleven psychological functions in terms of objects, states, movements, communications, and the like; and he concludes that "once carried through, the central description and derivation of the psychological functions should supply a sound basis for all the special psychologies" (p. 114).

Boring's "Psychology for Eclectics" manages to embrace the principles of structuralism, behaviorism, functionalism, and *Gestalt*-psychology, by accepting "structured wholes built upon both behavioral and phenomenal terms" (p. 124). "Epistemologically it is plain," writes Boring, "that a fact is a relation and that a relation is a function. . . . This conclusion leads to the rather surprising dictum that psychology deals only with meanings, for a meaning is just such a relation as we have considered a fact to be" (p. 125).

J. P. Nafe, in his article on "Structural Psychology," attempts to show that, despite the apparent systematic differences among them, psychologists nevertheless "form a homogeneous group whose interests, problems, and methods are similar" (p. 139).

Köhler's article "Some Tasks of Gestalt Psychology," deals primarily with group-formation and its physiological correlate.

In "Some Problems of Space Perception," Koffka remarks that "space-perception in all three dimensions is the result of organized brain-activity and . . . we can understand our space-perception only in terms of organization, i.e., in terms of actual dynamic processes, and not in terms of mere stimulus-sensation correlations" (p. 185). Perception is "the result of ever-changing stresses producing new and ever new organizations."

"Structure, Totality of Experience, and Gestalt," by F. Sander, gives the genetic emphasis to *Gestalt*-Psychology which characterizes the theories of Felix Krüger and the Leipzig psychologists. Structure, we are informed is not synonymous with *Gestalt*, but "denotes the psycho-dispositional constants conditioning the *Gestalten* of experience" (p. 192). "Every formation is experienced as a satisfactory fulfilment of some inner urge, possessing the whole consciousness with dull compressed feelings, an urge for formation of the formless, significance of the meaningless" (p. 194). "The specific directions of the separate sub-structures almost never chime together in an organic unity, though they are always borne within a whole, but usually in a high-strung whole, in which now the one, now the other, determines

the actual process of experience, attitude and action, though always in conformity to the immanent plan of the whole" (p. 203).

Pavlov in his "Brief Outline of the Higher Nervous Activity," notes that the anatomical substratum of the unconditioned activities is to be found in the sub-cortical centers. "When the sub-cortical center of the food-reflex is excited, all other stimuli reaching simultaneously the finest receptors of the hemispheres seem to be directed toward that center (whether directly or indirectly), and may become firmly connected with it. Then takes place what we have called a conditioned reflex" (p. 209). "The conditioned reflex may serve as an excellent object for the study of the nature of the individual cortical cells, as well as of the processes taking place in the whole cortical cellular mass, since the excitation of the cells of the cortex of the cerebral hemispheres serves as an initial stimulus for the conditioned reflex" (p. 210). However, the spreading of inhibition all over the hemisphere, lasting for a considerable length of time, when it is produced by a definite agent at a definite point, still remains incomprehensible. Is it due to a defect, or the inertia of the apparatus, or is it a definite phenomenon, the biological meaning of which still escapes us (which, of course, is quite possible)?

"As a result of the indicated work," continues Pavlov, "the cortex presents a grandiose mosaic, upon which are distributed at a given moment, a huge number of points of application of external excitations, either stimulating or inhibiting the various activities of the organism. Since, however, these points are in a definite, mutual, functional relationship, the cerebral hemispheres are at the same time, every single moment, a system in a state of dynamic equilibrium, which one might call a stereotype" (p. 212-213).

"In conclusion, it may be said that our experiments are, of course, only the first tentative experimental approach to one of the most important physiological questions of the interaction of the cortex and the nearest sub-cortical centers" (p. 220).

Schniermann's discussion of "Bekhterev's Reflexological School" reveals the difference between the conditioned reflex of Pavlov and the teachings of "reflexology." According to Bekhterev, reflexes are associated not only in the receptor but also in the effector part of the arc. The sleep-mechanism, for instance, is formed sub-cortically before the cortex is ready to function, and is therefore not attributable to a diffused inhibition of cortical processes as Pavlov maintains. "Reflexology has little in common with the foreign tendencies in psychology leaning upon subjective conceptions. American behaviorism (Anthroponomy) stands nearer to Bekhterev's reflexology, aiming at a strictly objective study of behavior, and also utilizing the evolutionary genetic method, though differences in the very 'method of knowledge' still remain" (p. 239).

Kornilov's "Psychology in the Light of Dialectic Materialism," is an application of the philosophies of Marx and Engels, and a revision in

terms of *Gestalt*. Thus Engels is quoted with approval upon the principles of psychological interpretation as follows: "How does the mind acquire these principles? Does it find them in itself? No—we deal with the form of existence, with the form of the external world, and these forms thought can in no wise draw from itself, but only from the outside world. Principles prove to be not starting-points, but are abstracted from them. It is not nature and human life which are guided by principles, but principles themselves are right only insofar as they agree with nature and history. This is the only materialistic interpretation of the question" (p. 260). Thus, Psychology is described as a unity of the subjective and objective, and thinking is as much a property of matter as is motion. Quantity is transformed into quality by a "leaping development," and the development of any phenomenon or system is always a *self-development*. "Whatever takes place in any part of the whole is determined by the internal nature of the structures of this whole" (p. 265).

Hunter, in his paper on "Anthroponomy and Psychology," defends the former as a science because "it does not have the limitations of its field set by philosophy" (p. 299). The author appears to take it for granted that philosophy and science are as irreconcilable as theology and science are sometimes thought to be.

The "Bio-Social Standpoint in Psychology," described by the late Albert P. Weiss, is a brief summary of the developmental history of man from a point of view wherein "the so-called mental categories are absorbed in the ontogenetic and phylogenetic analysis of bio-physical reactions and bio-social responses" (p. 306).

Dunlap's "Response Psychology" takes the view that mental heredity is "entirely outside the brain." All mental differences come from the periphery. The sole function of the brain is that of "integration." The "stock emotion-names . . . are really the names of typical stimulus patterns, and not the names of typical emotional expression, nor names of typical emotions" (p. 319). *Appet* is suggested "as the concrete term to designate an actual affective basis of desire" (p. 320). *Appets* are neither "psychic forces" nor "drives"; but experientible facts "just as colors, sounds, and other sentienda are experientible . . . Their being experienced depends on the excitation of certain visceral receptors, just as the experience of colors depends on the excitation of visual receptors. I have brought *appets* out into the periphery" (p. 320).

Woodworth finds that his "Dynamic Psychology" has room for both *Gestalt* and sensory analysis, while Spearman's "G and After—A School to End Schools," adds *p*, *o*, and *w* "broad factors of perseveration, orexis (striving), and will to the already well-known *g* and *s*."

Janet's "Psychological Analysis" reviews the clinical mind as a group of functions, not all of the same value, existing in a hierarchy maintained by psychological tensions, and "Psychoanalysis, its Status and Promise" is discussed by J. C. Flugel.

Adler's article on "Individual Psychology" restates the theory of inferiority. "From the sense of female inferiority, which most people, men and women alike, possess, both sexes have derived an overstrained desire for masculinity, a superiority complex, which is often extremely harmful, a will to conquer all difficulties of life in the masculine fashion, which I have called the *masculine protest*" (p. 398). One might suppose that such a protest would be feminine, and such a "complex" one of "inferiority," but the strength of "individual psychology" does not appear to rest in its logic.

Dewey's "Conduct and Experience" is an acute discussion of behavior and its implications for psychology. Some behaviorists have said that the psychologist uses perception, thought, consciousness, just as any other scientist does. To admit this and then go on to say (and act upon the saying) that, while they form no part of the subject-matter of physicist and physiologist, they do form a large part of the subject-matter that sets the problems of the psychologist, seems strange to me—so strange as to suggest an emotional complex" (p. 416). "To substitute linguistic behavior for the quality of acts that renders them 'mental' is an evasion" (p. 416). "The psychologist is concerned exclusively with experiencing. . . Experiencing has no existence apart from the subject-matter experienced; we perceive objects, veridical or illusory, not percepts; we remember events and not memories; we think topics and subjects, not thoughts; we love persons, not loves . . ." (p. 417). "A color or sound may be an object of an act termed sensing, and a tree or orange may be an object of the act of perceiving, but they are not sensations or perceptions except by a figure of speech" (p. 420). "If the acts of sensing, perceiving, loving, admiring, etc., are termed mental it is not because they are intrinsically psychic processes but because of something characteristic which they effect, something different from that produced by acts of locomotion or digestion" (p. 421).

In Kelley's paper on "The Inheritance of Mental Traits," I find both the English and the mathematics too difficult for comprehension.

Spearman's second paper in this volume on "Normality" is an attempt to escape the "danger of slipping overboard into the unfathomable seas of epistemology and metaphysics" (p. 444) by recourse to statistics with their "central values" and "dispersions."

The concluding essay on "Motivational Psychology" by Troland is a peripheral theory of response as a series of events which "may be considered as beginning with an object, or set of objects, in the environment of the organism" (p. 462). "This gives us a general basis for explaining nearly all types of affective experiences in terms of their afferent conditions" (p. 472). "The choice of alternative lines of conduct in the force of a given stimulus will be determined by the *greatest past affection*, which is proportional (according to our hypothesis) to the greatest present conductance. The doctrine, as a whole, is therefore *hedonistic* in character, but comprises a 'hedonism of the past' rather than of the present or the future" (p. 478).

Here, at length, is an indication of agreement with one of the previous authors, Dunlap, who has "brought appetts out into the periphery." One is reminded, however, of the experimental studies of G. E. Coghill and C. M. Child, which lead to quite opposite conclusions regarding the origin of motivation.

In the second volume under review, we have, for the most part, authors' summaries of what they regard to be their chief contributions to psychology. Thus, Baldwin, Calkins, Janet, Jastrow, Spearman, Stern, Stumpf, Ziehen, and Zwaardemaker, are content to give resumés of their work, with brief reference to their own birth, early life, teachers, and other influences.

It is remarkable how little "personality" creeps into these pages. The one notable exception is McDougall's essay with its unsparing frankness. It has been said that no one can fail to be interesting when he writes of himself. McDougall not only interests us, but also achieves literary form and poignant effects. "Rebellion" he writes, "illustrates a tendency of my nature which has, I think, played a principal part in determining my lines of thought and work. It is allied to, but not wholly to be identified with, the arrogance which I have already mentioned. Whenever I have found a theory widely accepted in the scientific world, and especially when it has acquired something of the nature of a popular dogma among scientists, I have found myself repelled into skepticism. This tendency had already led me to espouse the cause of psychophysical interaction, as against the then popular and orthodox parallelism and epiphenomenalism. Now it led me to active rebellion against the dominant theories of Hering" (p. 204). "In this spirit of defiance I wrote my *Body and Mind*, and gave it defiantly the sub-title, *A History and Defense of Animism* (1911). This, perhaps, is the most accentuated illustration of that uncompromising arrogance which I have already mentioned" (p. 209). "My *Social Psychology* had been meant, not as an introduction to the field, but rather as an indispensable preparation or propaedeutic. *The Group Mind* was a part of my projected *magnum opus*; but its reception was so unfavorable that the *magnum opus* went a-glimmering. For, as I have said, I have found it increasingly difficult to believe in the value of my work" (p. 212).

There are personal touches in Claparède, along with a modest account of problems and achievements. Dodge reveals himself in the statement of his scientific aim, which "may be expressed as a persistent effort to record with accuracy the behavior of normal and abnormal human organisms at various levels of neural integration, and to describe and understand that behavior as to its conditions, its variations, and its modifications as the various levels interact to produce overt acts" (p. 121).

Kiesow's essay is remarkable in its avoidance of any significant reference to his own life and achievements. He does give an interesting picture of the early days of Wundt's laboratory, and especially of the effect upon Wundt of the dissident views expressed by Külpe in his *Grundzüge*.

"Carl, you are always getting into things." This oft expressed opinion of his mother, is the key to Seashore's account of his life and works. It is a "success story," modestly told, but deals more with the institutional history of psychology in the State University of Iowa than with the advancement of psychology as a science.

Warren's account of his own progress "from mysticism to mechanism," work is to a degree self-revealing, and contains some interesting historical data, but the progress of psychology as a science is seen in its external rather than in its internal conflicts.

It should perhaps be added that the essays of Stern and Stumpf are translated from a German publication, *Die Philosophie der Gegenwart in Selbstdarstellungen*. While these two are more "philosophical" than the other essays which make up the volume before us, there is little or no reference to experimental results in any paper, excepting that of Zwaardemaker, the physiologist.

In this book, as in the *Psychologies*, one receives the impression that the psychologist is more interested in his theories than in facts. Though he professes disdain for epistemology and metaphysics, he is in the main groping for a theory of knowledge that will provide him with a sound basis for his psychological methods.

We need a psychologist with the calibre of the late Oswald Külpe. He did not blind himself to the epistemological nature of all scientific methodology; he stoutly maintained that without first reaching some agreement on principles of method, there could be no "scientific" results in psychology.

R. M. OGDEN

Cornell University
Ithaca, New York

KARPF, M. J. The scientific basis of social work. New York: Columbia Univ. Press, 1931. Pp. xx+424. \$3.75.

THOMPSON, D. A mind that was different. Oklahoma City, Okla.: Harlow, 1931. Pp. xvi+117.

TRAVIS, L. E. Speech pathology. New York: Appleton. Pp. xxxiv+331. \$4.00.

PROBLEMS OF MEASURING CHARACTER AND PERSONALITY*

From Yale University

MARK A. MAY

All scientific enterprises of major importance are, at different stages in their development, confronted with at least three types of problems. First, are the theoretical problems which usually center around the question of whether or not the enterprise is at all possible. Second, are questions of technique. If the job is possible what techniques are necessary for its accomplishment? Third, are questions of practice. The techniques may be available and the enterprise may be realized, as, for example, in television, but it is not commercially practicable.

For the purposes of our present discussion the problems involved in the measurement of character and personality may be divided into the theoretical, the technical, and the practical. While no attempt will be made here to solve any of them, yet ways of approach will be suggested.

I shall not dwell long on the theoretical problems because I suspect that the greater interest is in the technical and practical questions. The only excuse for discussing the theoretical issues at all is that they are really basic to an understanding of the practical and technical ones.

The following syllogism is familiar to all students of measurements:

All that exists exists in amounts.

Any amount can be measured.

Therefore anything that exists can be measured.

Character and personality exist, they exist in amounts, hence they can be measured. Without stopping to examine the fallacies in this theorem, let us assume that character and personality can be measured in one way or another, and turn our attention to the ways in which it may be accomplished.

The theoretical issue is whether or not it is possible to measure character and personality, or, indeed, any mental or social variable, with the same objectivity and precision as objects of the physical

*Received in the Editorial Office, March 5, 1931.

world are measured? If not, then is the difference merely one of degree, or is it a fundamental difference in kind? There are those who say that it is only a difference of degree, that mental and social measurements are fundamentally the same as physical measurements the only difference being that of precision. Others take the position that the difference is much more fundamental and that mental and social measurements must by their very nature be of a different order than those of the physical world. I shall not be lured into this controversy because it leads into the philosophy of measurement, a topic which I am incompetent to discuss. I shall, however, venture the statement that there is no great gulf fixed between the data of the physical sciences and those of the social sciences.

The distinguishing features of most physical measurement are the properties of objectivity, consistency, equality of units, and a defined point of reference or zero. Mental and social measurements approach those of the physical sciences in the degree that these attributes are present.

The theoretical question stated above may now be split into three sub-questions. The first is to what extent is it possible to achieve objectivity in the measurement of personality? Second, to what extent are the traits, qualities or attributes fixed and stable objects which can be measured in a consistent manner? In other words, how can we know that the object of measurement is not itself changing either qualitatively or quantitatively or both? Third, to what extent can character and personality qualities be measured in equal units which are counted from a definite point or reference?

If it is true that most character and personality tests rate low in respect to objectivity, consistency and equality of units, what are the prospects of improvement? Will technical improvements raise them to a higher level? Not unless the theory on which they are built is sound. It is my opinion that this theory is not sound and that no amount of technical improvement will suffice.

Most measures of character and personality assume a trait or a type theory. If personality is an aggregate or even an integration of traits, then it follows that these traits should be identified and measured. Each trait is supposed to be an independent entity as far as measurement is concerned. It is supposed to exist in amounts in the sense that there are quantitative individual differences in it. Thus the theory assumes that traits exist in entities and are therefore measurable.

Suppose we deny that traits exist, can we then reverse the theorem and say: that which does not exist cannot be measured. This leads us to the brink of a terrific metaphysical abyss—to wit, the nature of the non-existent. Keeping away from this chasm, we need only to point out that the possibilities of measurement of character and personality depend to a great degree on the view we take of what constitutes a trait. My own view is that traits are only convenient names given to types or qualities of behavior which have elements in common. They are not psychological entities but rather categories for the classification of habits.

If this view of the nature of personality is taken, then the problem of measurement is not one of selecting a trait as an entity and proceeding to construct an instrument for measuring its quantitative variations. Practically all the data of character and personality studies are in reality only so many samples of different types of responses. The same is true of intelligence and educational achievement test scores. A raw test score does not tell us *how much* of a given trait or quality an individual possesses, it tells us only *how many* responses of a certain kind that individual made under the conditions imposed by the test. It is true, however, that the total raw scores on some tests are capable of being transmuted into a scale of equal units so that the modified scores may be treated as though they represented an amount of something. But this is a secondary and statistical process. The raw data of all psychological tests are nothing more nor less than enumerations of responses. An honesty test is a device for recording samples of honest or dishonest behavior; an emotional instability test is a device for securing symptoms or samples of emotional responses; an attitude test secures sample opinions, or preferences. The same may be said of all varieties of the so-called character and personality tests.

If this view of the matter is taken then personality testing is not only theoretically possible but also practical. When the testing program becomes a sampling program it meets the scientific requirements of objectivity, consistency, equality of steps, and the technical problems of reliability and validity are greatly simplified. No assumptions are made concerning the nature of what is being measured because we are not measuring psychological entities but merely sampling behavior. There need be no confusion concerning what kind of behavior any response is a sample of, for it is a sample of that kind of

behavior with which it has elements in common and these common elements are determined empirically.

We come now to the technical questions as to how this sampling process can be carried out. What kinds of samples are needed and how may they be obtained? Consider, first, the problem of defining the area or type of behavior that is to be sampled. In mental and educational testing this is an easier task than in personality testing. Such areas of knowledge as arithmetic, spelling, reading, and the like are easier defined than behavior qualities such as aggressiveness, caution, deceit, and tact. It is here that the criterion of objectivity must be fulfilled. It is not enough to say that we are going to sample aggressiveness because there is too much disagreement as to what types of behavior would be regarded as aggressive. It is doubtful whether any definition of aggressiveness could ever be framed in behavioral terms to which all would agree. One procedure is to define behavior in terms of types of responses to specific situations. We can say that to any situation there are a group of responses which have in them an element which may be defined as aggression. Another plan would be to collect a group of responses to a type of situation and classify them in a variety of ways. One classification would be according to one type of common elements and at the same time these same responses might be grouped in other ways.

A second technical problem is that of selecting an appropriate recording device. The problem is that of securing accurate and complete records of what individuals actually do in situations of daily life or in controlled situations of the laboratory. There are a great many different ways of doing this. I shall mention only a few of the more common ones.

The first is the method of observation. This method has been used widely and successfully in securing records of animal behavior, but, surprising as it may seem, it has been used very little in recording samples of human behavior. To observe what a child or an adult does in any situation would seem to be a very simple and easy task, but, as a matter of fact, it is very complex and difficult. One difficulty is that of devising suitable shorthand or notation of recording. Another difficulty is that of knowing what to observe. Those who have experimented with the method are of the opinion that most observers are incapable of keeping track of more than one or two features of a complex act. But techniques of observing and of re-

cording observations are being developed rapidly, and this method promises to be one of the most prominent and most useful.

A second method is that of testing. I have already called attention to the fact that most raw test scores are samples of responses to test situations. I wish to emphasize the fact that all behavior or performance tests and nearly all paper and pencil tests are devices for securing systematic and formal records of behavior. An honesty test, for example, is a device by which the child leaves a record of his honest or dishonest behavior. The difference between observation and testing is that in observation the observer records the behavior, while in testing the subject records it. The record is the test score and the score is a quantitative description of a sample of behavior. Thus the test measures only in the sense that it samples.

A third way of securing records of behavior is by the use of laboratory instruments. Speed of reaction is recorded on a clock or a smoked drum; muscular fatigue is recorded by an ergograph; emotional responses by a galvanometer; and so on for other types of behavior. But the varieties of behavior that can be recorded with instruments of this sort are limited. Yet, on the other hand, this method has the advantage of being very precise and the samples of behavior thus secured may be described more truly in quantitative terms.

Other methods of securing records of human behavior, opinions, attitudes, and abilities are the questionnaire, the interview, the case history, and the autobiography. In each of these the record is descriptive and qualitative but has the advantage of being more complete. The rating method has the appearance of being quantitative but, as a matter of fact, it is not. Numerical ratings are only number-labels placed on opinions. They are nevertheless valuable as condensed reports of behavior provided they are based on observations or other records of concrete acts.

No rule can be given for the selection of a method for the recording of behavior. Much depends on the nature of the behavior and the purpose of the study. For recording relatively simple and specialized types of response such as speed of reaction, some laboratory instrument should be used; but for recording a complex act such as a quarrel, fight, or temper tantrum cruder devices must be used. In general, one should strive to make all records of behavior as precise, as systematic, and as quantitative as possible.

A third technical problem is that of determining the reliability of

the method of recording. It is a well-known fact that most character and personality tests have low reliabilities. This is due in part to the statistical procedure commonly employed for finding reliability. The reliability of a test, or of any instrument of measurement, is usually determined by finding the self-correlation between two applications of the same test or of two similar forms of a test. There are two major difficulties with this procedure. The first is that there are certain mathematical assumptions underlying it which are rarely, if ever, completely fulfilled by most test scores; the second is that the behavior, or attitude, or whatever is being measured, is not a constant in any individual. We cannot assume that the individual possesses a certain amount of honesty, or courage, or tact, or even intelligence, and that this amount is relatively constant even over short ranges of time.

Suppose, for example, the aggressiveness of an individual has been sampled a large number of times in the same situation, or in situations as near the same as it is possible to get them. There will undoubtedly be considerable variability among these samples. Part of this variability may be attributed to variations in external conditions and in the procedure of recording the behavior; but part of it is due to the fact that the individual himself varies. So there are two kinds of variation in the test scores of a single individual, variation in the procedure and variation within the individual. It is important to keep them separate. The reliability of the test is dependent on the first alone. Self-correlation between two sets of scores gives an estimate of the average amount of individual variability including both kinds. What is needed is a technique for estimating the amount of variability in an individual's scores that may be attributed to variations in testing or sampling procedure, apart from variations within the individual.

A fourth technical problem is that of validity. The question of validity is usually stated thus: Does the test measure what it claims to measure? Back of this question lurks the assumption of measuring entities. It is assumed that there is an entity of intelligence, or of mathematical ability, or of honesty, and the like, and the validity of the instrument is how adequately it measures the amount that any individual possesses. As soon as we give up the notion of measuring entities and take on the idea of sampling behavior, the question of validity is somewhat altered. There are really two important questions. The first is, how pure are the samples? How well do they

fulfill the original definition of the behavior that is being sampled? If it is aggressiveness that is being sampled, the question is, to what extent is aggressiveness as defined represented in the sample? What else besides aggressiveness is also in the sample? If there are other factors or elements in the sample, how may they be eliminated? The second question is how many samples are necessary to cover the ground; or, how many are necessary before we can say that that type of behavior has been adequately sampled?

The traditional method of validating a test by correlating it with some "outside" criterion is of very little use. This procedure is based on the assumption that a test measures best that with which it correlates the highest. This false assumption has caused test makers no end of trouble. The purity, and hence the validity, of any sample of behavior is determined not by finding how much it resembles something else, but rather by how much it resembles another sample of the same trait, or, better still, by how well it squares with the original definition of that behavior.

There are a number of validating techniques all of which have as their central idea that of correlation or comparison. It has been claimed that the best process of validation is to compare the subject's tests scores with a clinical picture of his personality, or with a case history, or with a psychiatric analysis. Actual coefficients of correlation are not used, yet the test scores are compared with a different measure of presumably the same trait.

All these techniques assume that the trait or behavior quality is an entity and that we are seeking one single linear measure of it. If, on the other hand, we make no such assumption and regard our test scores as samples, the purity of which we wish to determine, the process of validation really becomes one of determining internal consistency. By internal consistency we mean consistency of the samples with each other and with the original definition of behavior. One way to determine this internal consistency among samples is to collect them by different methods. For example, suppose we are interested in deception. We might collect samples by the testing method, by the laboratory method, by the observation method, and even by the interviews, all on the same type of situation. Thus, by varying the technique, using the same subject in the same situation, we would get data that would reveal the internal consistency of our samples. The first method, therefore, of purifying the samples or of determining their impurities is by correlations between samples collected

by different techniques. A second method is that of correlating one sample with a total of many. This is akin to the old and discarded technique of correlating subtests with total score. But the method is entirely justified, provided certain assumptions are fulfilled. One is that the impurities in the samples are uncorrelated with each other so that when a large number of samples are lumped together the irrelevant factors will cancel out and the total or average score will be at least theoretically pure. If, however, the number of samples is small, a third method is possible. The basic datum is the average intercorrelation of the samples. The procedure is to find the average inter- r of the samples, substitute it in the Spearman-Brown formula, and take the square root of the result. This figure is the predicted correlation between the total number of available samples and an infinite number of the same kind.

A fourth validating technique is the Spearman tetrad method. If the tetrads sum to zero with a standard deviation less than chance would provide, then we know that there is a common element among the samples.

A fifth method of validating or determining the purity of the samples is by proper scoring devices. But this is a technical question which deserves consideration on its own account and will be considered later.

We may turn now to the second major question of validity, which is the adequacy of the sample. How do we know when we have enough samples, or when we have covered the area of behavior? In how many different types of situations should the behavior be observed or tested before a reasonable prediction can be made that the individual will show this behavior in any situation permitting it? No numerical answer can be given to this question. Much depends on the type of behavior and on the individual. Statistically, it is a matter of the spread or variance in the individual. There may be so much variation in the behavior of an individual that reasonable prediction for him is out of the question. The first thing is to recognize individual difference in variability in the same behavior. There are some individuals who show small variability, and for them a small number of samples will suffice. There are others with much greater variability for whom many more samples are necessary. Ideally, we should continue to sample until the variance of each individual becomes stable or nearly so. The more samples we have for an indi-

vidual up to a certain point the greater will be his variance. When his variance ceases to increase then we can quit sampling.

There are at least two ways of determining the adequacy of a sample. If it is possible to secure a large number of samples for each individual, then individual variance may be determined. This variance will increase as the number of samples are increased up to a certain point provided the samples are random. The rule should be to continue sampling until the variance ceases to increase. There will be individual differences in the number of samples required.

If, on the other hand, only a few samples are taken for each individual, it is possible to predict how many more will be required. The procedure is to find the average intercorrelation between the obtained samples and, by substituting this figure in the appropriate formula, a prediction may be made concerning how many samples are necessary to yield a correlation of .90 or .95 or any other desired amount between the total number of obtained samples and an infinite number. All this is based on the assumption that the obtained average intercorrelation will not change as the number of samples is increased.

A fifth technical problem is that of scoring. Test makers in the field of character and personality have paid too little attention to the development of adequate scoring plans. Many of the preceding technical problems such as validity, reliability, objectivity, and equality of units depend in no small measure on the way the test is scored. In fact, the usual criteria of scoring are stated in these terms. The general rule is to score all tests, and all other records of conduct in such a way that the results will be as quantitative, as reliable and valid as possible.

Consider the problem of scoring for high reliability and validity. In the light of the foregoing, the chief problem in testing is that of securing an adequate number of pure samples of the behavior that is being studied. One of the surest ways of purifying the samples is to score out the impurities. A pure sample is one which contains a complete description of the behavior in question and nothing else. An impure sample is one which contains a partial description of the behavior in question plus other things. Theoretically pure samples are seldom if ever secured, but they may be approximated by careful and skillful scoring.

The problem of quantification of samples is by far the most difficult of all problems thus far mentioned. The data of character and personality studies vary all the way from qualitative descriptions

such as case histories, autobiographies, results of interviews, to records that are in equal units such as measures of speed in time units. Most test data are in a crude semi-quantitative form. They are for the most part enumerations or counts of responses. The scoring problem is one of quantifying these records in ways that will conserve their meaning.

I am convinced that the only way to handle this problem is to make all records or samples of behavior in ways that can be counted. Never mind whether the units are equal or not, but be sure that all records are in terms of things that can be counted. No progress can be made in the direction of quantification until counting is possible. For example, if we are interested in truthfulness the first and most important thing to know is how often the subject lies. For the purpose of understanding character the frequency of lying is much more important than the magnitude of the lies, which, after all, is a matter of opinion.

The second step is to record the samples so that they can be arranged in some order. If we have 25 sample lies told by John Doe and if these can be arranged in an order of seriousness or importance, we are that much nearer to a quantitative scale. For, after all, the first thing we want to know about John Doe is the probability that he will lie in a given situation and, in the second place, how big a lie he is likely to tell. The measure of his truthfulness is the probability of his veracity.

When it comes to the question of equality of units I am convinced that equal units in personality testing are possible only in a statistical sense. I do not believe that there are such things as equal amounts of honesty or courage or tact in the sense that each of these is some sort of a psychological continuum of which each individual possesses a certain amount. The only sense in which two individuals are equally honest is that they are each equally likely to behave honestly in a given situation, or that the proportion of situations in which they are honest is the same. The advantage of transmuting raw scores into sigma scores, percentile scores, or T-scores, or any kinds of statistical units, is not that it will enable us to say that one subject is twice as tactful or twice as cautious, or one-half as honest as another, but that it affords a ready and convenient means of comparing the scores of one test with those of another.

When we deal with large numbers of samples it is quite possible that their totals behave in quantitative fashions. We have repeatedly

asserted that a total score on a test is simply a count or enumeration of correct responses which is a sum of unequal and incomparable samples. If individual A gets a score of 120 on a test, this means that he has 120 correct responses. B may also have a score of 120 but not the same 120 items, hence their two scores are not the same. If A has ten coins and B has ten coins it does not follow that they each have the same amount of money. But if they each have 1000 coins selected at random from all the possible kinds of coins, then the chances are much greater that each has the same amount of money. It is reported that Sears, Roebuck and Company once estimated the dollars' worth of orders for a day by the simple process of weighing the mail.

By the laws of permutations and combinations it is possible to show, I think, that scales of equal units may be constructed from raw data which are only enumerations of responses. If enough samples are taken and if reasonable care is exercised in getting them as pure as possible, it is quite likely that two numerically equal total scores may mean exactly the same thing in the sense that they represent the same probability of occurrence of the behavior in question.

A sixth and final technical problem is that of norms. Most of current character and personality tests are without norms. There are several reasons why this is so. In the first place, the authors of these tests have not followed through with enough applications of their tests to secure norms. A second reason is that satisfactory bases for norms are difficult to secure. In most intelligence tests chronological age is the basis of norms; in educational achievement tests it is either age or grade, sometimes both. But age has not been satisfactory as a basis for norms in personality tests for the reason that it does not correlate high enough with test scores. While it is probably true that certain types of behavior, attitudes, and skills increase or change with age, these changes are also tied up with other variables in such ways as to reduce what might otherwise be a high correlation with age.

A norm is a statistical average which constitutes a kind of a standard of expectation. If a boy is ten years old he is expected to be in the fifth grade, and have a mental age of ten on the Stanford-Binet test. In the realm of behavior certain norms or standards are prescribed by common sense based on crude observations. For example, children are supposed to pass through a stage when it is "normal" for them to exaggerate or tell lies of considerable magni-

tude. This behavior has been noted in many children between the ages of three and five. Parents and others tend to regard such fanciful exaggerations as normal behavior belonging to that stage of development. Other examples of types of behavior that seem to belong to different stages of development are easy to find. They illustrate the fact that there are types of behavior which presumably change with age and in a more or less orderly fashion.

Very little research has been done in the direction of defining these standards of expectation. When the matter is investigated there may be fewer of them possible than we now expect. Standards that are appropriate for a heterogeneous population are not easily discovered. Character and personality tests require for their adequate interpretation not only population norms but also individual norms. That behavior which is normal for one individual is not for another. This comes back to the problem discussed in the section on validation where the need for more data on individual variability was pointed out. Abnormal behavior may be deviations from the group standard or from a standard that applies only to the individual. The problem of norms in personality testing requires further analysis of what is meant by normal and abnormal behavior, and also further research on the individual variability.

We come finally to the practical problems involved in personality and character testing. Lists of practical considerations have been prepared by Goodwin, Watson, Hartshorne and May, and others. I shall repeat here some of these and will add others.

There are in existence no less than two hundred and fifty character and personality tests or testing techniques, one hundred at least are published and can be purchased. The first question then is how to select. What is needed is a kind of an accumulative encyclopedia of personality tests giving a few of the more pertinent and practical facts about each. No one can give off-hand a prescription as to which tests are more valuable, for too much depends on local situations.

This raises a second question. The total area of personality is so much wider and covers so many more qualities than is covered in the whole field of educational testing that it is quite out of the question to test the total personalities of school children in the short and simple ways that we are accustomed to in testing intelligence and educational achievement. An adequate character testing program

must extend over a period of not less than six weeks or ten weeks and will require at least one test or series of observations each day.

A third practical difficulty is that more skill is required to administer personality and character tests than is required to administer an intelligence test. In the case of intelligence tests the instrument becomes standardized and any teacher can apply and score it. This is not the case with most personality tests. Here the technique is the important thing, and the teacher or tester must first master it. This is especially true of the behavior or performance type of test.

A fourth practical problem is the use of short cuts. Most schools use ratings as a quick and easy way and have used them for a long time. I am convinced that this is not a bad thing provided improved rating techniques are available. If teachers would work out a plan or system of observation of the behavior, attitudes, etc., of their pupils and agree on a simple but uniform method of recording their observations, then later, on the basis of these observations rate their pupils, the results would be very worth while. These ratings could also be checked by a few standardized tests. But as long as ratings are in the realm of guessing and prejudice they are worth very little. Good use can be made of case histories and other behavior documents, provided they are systematically done. Diaries and autobiographies and other similar records are also useful. Teachers could be of great service to science if they could encourage children to keep regular and fairly detailed diaries.

The main thesis in this paper has been that most of technical and practical problems in this field are unnecessarily complicated by assuming a trait theory of character. I have shown in some detail how the technical problems of test making are needlessly involved because of this theory. The same is true of many practical problems. As long as school principals, superintendents and other educational administrators assume that growth in character is a matter of adding to or otherwise expanding a series of traits, and that there should be devices for measuring these changes, they are doomed to disappointment. On the other hand, the collection of samples of behavior, attitudes, and opinions of school children is much simpler. In all this work the guiding formula should be that personality can be most easily and most scientifically measured by taking samples of behavior.

Yale University
New Haven, Connecticut

LES PROBLÈMES DE LA MESURE DU CARACTÈRE ET DE LA PERSONNALITÉ

(Résumé)

Pour rendre plus facile l'étude des problèmes de la mesure du caractère et de la personnalité, on les a divisés en trois groupes—théoriques, techniques, et pratiques. Le problème théorique le plus important est le suivant: est-ce qu'on peut employer les mêmes méthodes pour mesurer le caractère et la personnalité que pour mesurer les objets du monde physique? La possibilité de la mesure dans toutes les sciences biologiques dépend beaucoup de la théorie qu'on soutient sur la nature des phénomènes de l'observation. Si on suppose que les traits de caractère et de personnalité soient des entités psychologiques qui se trouvent chez les individus et dont les individus possèdent des quantités différentes et qui se manifestent en types de comportement, nul essai de mesurer ces entités métaphysiques ne réussira, parce que ces entités n'existent pas. Si cependant on considère les données des sciences biologiques comme échantillons du comportement, et que la mesure se compose de compter et d'énumérer ces échantillons, la mesure devient possible. Les résultats totaux d'un test ne nous disent pas combien d'un trait ou d'une qualité donnés l'individu possède; ils nous disent seulement combien de réponses d'un certain type l'individu a faites dans les conditions imposées par le test.

Les problèmes techniques considérés sont: (1) celui de diviser les types de comportement dont on peut facilement isoler des échantillons; (2) celui de choisir de bons appareils à les noter; (3) celui de déterminer la vraie valeur exacte de ces appareils; (4) le problème de validité; (5) le problème d'évaluation.

Les problèmes pratiques considérés sont ceux des types des tests qui existent actuellement, de leur standardisation, du nombre nécessaire pour obtenir une mesure de valeur, et enfin des types d'adresse nécessaires dans l'administration et l'évaluation de ces tests.

MAY

PROBLEME DER CHARAKTER- UND PERSÖNLICHKEITSMESSUNG

(Referat)

Die Probleme der Charakter- und Persönlichkeitsmessung werden bequemerweise in drei Gruppen geteilt: theoretische, technische und praktische. Das wichtigste theoretische Problem ist die Frage, ob Charakter und Persönlichkeit wie Objekte der physischen Welt gemessen werden können. Der Umfang der Messbarkeit in allen biologischen Wissenschaften hängt zu einem beträchtlichen Grade von der Theorie ab, die man über die Natur der zu beobachtenden Erscheinungen aufrecht erhält. Wenn wir annehmen, dass Charakter- und Persönlichkeitszüge psychologische Wesenheiten sind, die den Individuen innewohnen, und worin sich die Individuen dem Inhalt nach unterscheiden, und die sich in Typen des Verhaltens manifestieren, dann müssen alle Versuche, diese metaphysischen Wesenheiten zu messen, scheitern. Diese Wesenheiten existieren aber nicht. Wenn wir jedoch die Aufzeichnungen der biologischen Wissenschaften als Proben des Verhaltens auffassen, und wenn die Messung darin besteht, dass man die Proben zählt oder nummeriert, dann ist die Messung möglich. Der Gesamtestkoeffizient gibt uns nicht an, wieviel eines gegebenen Zuges oder einer Eigenschaft ein

Individuum besitzt; er gibt nur an, wieviele Reaktionen einer gewissen Art das Individuum macht, unter den durch den Test auferlegten Bedingungen.

Die in Betracht gezogenen technischen Probleme sind: (1) das Teilen der Verhaltenstypen, die hinreichend genug Proben gestatten; (2) die Auswahl geeigneter Registrierapparate; (3) die Bestimmung der Zuverlässigkeit der Registrierapparate; (4) das Problem der Gültigkeit; (5) das Problem der Anschreibung (scoring).

Die in Betracht gezogenen praktischen Probleme sind: die jetzt zur Verfügung stehenden Tests, ihre Normierung, die Anzahl, die angewandt werden muss, um eine hinreichende Messung vorzunehmen, und endlich die Fertigungsarten, die notwendig sind zur Handhabung und zur Anschreibung dieser Tests.

MAY

THE FUNCTIONS OF CONVERSATION*

RAYMOND R. WILLOUGHBY

It is matter of common observation, though possibly not of common cogitation, that the great bulk of human speech has not yet evolved past a system of clichés and signals. At one end of the scale, for instance, it is difficult to see that "Howrya?" has any points of superiority to "Woof!" as a form of greeting, and it is even conceivable that the latter connotes a more sincere species solidarity; at the other, the invention of complicated modes of communication leaves the inventors aghast at the discovery that there is nothing of importance to communicate. And in the less utilitarian spheres, the most charming tangible exterior often masks a cold and rather horrible vacuity; as in the familiar situation portrayed by Walter Prichard Eaton in *Penguin Persons and Peppermints*, where the hostess kindly and graciously makes the young author at ease and starts him talking about his latest book, then fades gracefully away with a skill born of long practice, to repeat the maneuver on another unsuspecting victim.

There are a few attempts to reduce these matters to comprehensibility; the ordinary parliamentary rules of order may be regarded as an endeavor to get group thinking done by greatly curtailing the amount and originality of the contribution of each member of the group. E. B. South recently conducted a study of "committee work," using students as subjects in rather artificial situations, and focusing his attention, not on the mechanisms, but on the gross results as expressed in time and errors. The newer adult education movement (not that limited to teaching English or stenography in night schools by traditional methods, but the Ashland College approach has developed a renewed interest in the function of the leader as co-ordinator and the technique necessary to insure the maximum contribution from each member; of these the Inquiry studies on group thinking are the outstanding examples. All these are devoted largely to the external, immediately accessible aspects of the matter; the approach of Trigant Burrow and his group, though expressed in an excessively obscure manner in the voluminous writings of Burrow himself, are

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office.

evidently aimed at something more fundamental—namely, the deeper psychological mechanisms which find their surface expression in those traits of the group members which hinder (or more rarely facilitate) the attainment of maximum satisfaction from group inter-stimulation.

Believing the matter to be at once very fundamental and insufficiently examined, the writer undertook an exploratory study in the fall of 1928, designed to ascertain what motives and needs could be discerned in their own conversational behavior by a group of mature, intelligent, and somewhat psychologically minded men and women, including a fair proportion who were conspicuously successful conversationalists. The results have up to date baffled extended fruitful analysis, but it has been decided to publish a brief summary of them in recognition of their considerable interest and possible suggestive value.

The following was the form of inquiry used; the numbers were not in the original, but are inserted because several subjects answered the questions serially:

I should greatly appreciate from you some free associations on the subject of the *purposes of adult conversation*. My curiosity began by observing that the great bulk of conversations as they actually occur are peculiarly unsatisfying to all parties. The present inquiry is wholly exploratory, and whether anything farther can be done will depend on the nature of the requested very general reactions. Replies may be placed on this sheet; they should be long enough to be clear, but need not be longer than can be contained on both sides. The following suggestions should make plain the general nature of the comments desired:

(1) What are your ideals of social intercourse as mediated by speech?

(2) What psychological needs in you are satisfied by refreshing conversation?

(3) When conversations are satisfactory, what are the outstanding reasons?

(4) When unsatisfactory?

(5) Are there types of conversation which should be distinguished in such an inquiry?

(6) Types of participants?

(7) What is the effect of numbers on the satisfactoriness of conversations?

(8) Of topics?

(9) Should a distinction be made between ideal conversations and conversations which, though satisfying, are not ideal?

(10) Between the motives one would wish to animate him in conversation and those he finds actually do animate him?

The inquiry was sent to 25 men and 25 women, practically all known to the writer and for the most part knowing him; 20 were returned, not counting a few rambling remarks included in a personal letter to an associate of the investigator; the latter were by a young woman of about 20, a teacher in a western state, of good intelligence and well-poised personality, education equivalent to about two years in college; they were to the effect that she liked conversations like those A and B had (which may be characterized as rather intellectual discussions conducted largely on A's initiative, in attempts to develop his thinking on various topics) and like those C had when he felt in the mood (which may be characterized as sparkling, ephemeral camaraderie). Of the 20, one, from a male psychoanalyst and physician of middle age, merely stated that he was interested but unable to cooperate.

The smallest possible ratio of males to females returning usable answers was thus 3.0, i.e., 15 to 5; the males were also somewhat more prompt in returning them, on the average. Since nothing was done to weight either group for probable interest or lack of interest, and since it was anticipated that because women are customarily regarded both by themselves and men as the arbiters of conversation their ideals and practices would be rather clearly formulated, this result is of considerable interest. It is perhaps not quite fair to suppose that the attitude of a well-to-do society leader of middle age is representative: "Why, I couldn't answer that! I should have to sit down and *think* about it!" Nevertheless, it is reasonable to guess that women are much less impelled by any necessity to understand or direct their conduct in such matters—more "unconscious." They find themselves looked to to "make conversation," and they make it; they are bored or entertained at the result—anything deeper is not important, in fact, is perhaps pedantic. There are, of course, conspicuous exceptions; one of the most searching analyses received was from a young woman, and several of the male replies were decidedly superficial in nature.

Next above the level of not answering at all may be considered the "amusement" level. This is best formulated by a man of about 40, a teacher of psychology whose philosophy of social relations is defin-

itely Machiavellian; his intelligence, of course, is most superior, and his personality adjustments are good:

The ideal social intercourse is complete pleasure . . . there are only two motives, one for immediate pleasure, the other for

- (1) deferred pleasure. These two may come into conflict in any given conversation and baffle one somewhat. If one is completely bored, it is because neither motive is present.

A less sophisticated version of the same motive is that of a woman of about 25, a college graduate of the highest intelligence and excellent personality adjustments, employed as a psychological secretary:

My ideal of social intercourse as mediated by speech is en-

- (2) joyment. The most satisfactory conversations are those with congenial people, and vice versa.

The same motive, but somewhat overlaid by more penetrating analysis, runs through several other replies. Evidently there is very little that can be said about it, except that—and because—it is superficial; it is deeply colored by what passes as psychological thinking among the crowd—viz., the belief that a sufficient explanation has been given to a phenomenon when its name has been changed; for nothing seems clearer than that a satisfying conversation is satisfying, even though the satisfyingness be called pleasure or enjoyment.

There is also the conception that the function of conversation is intellectual—the conveying of ideas. The following is from a man of about 75, a classical scholar who has spent his life as a teacher of boys; his intelligence has, of course, been of the best; his personality has been a dominant one, and now shows evidences of incipient senility:

Conversation is to give and receive information, to secure refined and enriching human intercourse, and to share with

- (3) others the broadening facts of science. . . . Intelligence lies at the root of the whole matter and spirit of conversation, intelligence reinforced by knowledge and many contacts with life, of course.

A woman graduate student in psychology of about 25, of splendid personality and intelligence, combines the amusement and intellectual ideals, maintaining meanwhile that she has no ideals in such matters:

I have no "ideals" of social intercourse as mediated by conversation. The sole object of language is to convey thoughts,

- (4) and whether one is interested in steam engines or the number of hairs on the thorax of *musca domestica* it serves the same

purpose. Personally, I insist on being amused or interested—but it can at times be done by other methods!

This approach seems to the writer to abdicate most of the facts in favor of a loyal adhesion to the preconceptions of intellectualist psychology. Language may, of course, be used to convey thoughts, but such use seems exceptional, unless by "thoughts" is to be understood the whole gamut of feelings, urges, despairs, hopes, etc., which empirically are expressed by language. It is noteworthy that the intellectualist approach is relatively rare in the replies received, and even the lady mentioned is by no means consistent in her adherence to the intellectual ideal.

A third level may be discerned in the reply which avers that pleasure is the summum bonum of conversation, but endeavors to analyze the pleasure in terms of some form of expression, ego, or power satisfaction. This is usually to be gained at the expense of the other person, as by impressing him with one's own importance. The following is by a man of 23, a graduate student in psychology, of quiet and somewhat self-centered personality, aristocratic in manner, perfectly poised and of excellent intelligence:

- Refreshing conversation satisfies the expression of self tendency. . . . unsatisfactory when the ego is not allowed to be satisfied.

A man of about 35, a teacher in experimental schools and something of a social radical, a strong adherent of the psychology and philosophy of Stanley Hall, of excellent intelligence but somewhat unstable personality, contributes the following along the same line:

- Ego satisfaction of a negative sort—i.e., recognition of my ability . . . may be complete ego satisfaction—I am worth conversing with. . . . We think along like channels—I am normal.

A male graduate student in psychology, of excellent intelligence but rather unstable personality:

- I find it hard to give them an interested and sympathetic audience because of preoccupation with personal problems
(7) rather tend to portray things in a way that will give others an opinion of me and my past history such as I would like them to have.

A male psychoanalyst of about 35, English, student of Rank, now located in a western city, personality and intelligence excellent, includes in his list of tendencies satisfied:

Showing off power, giving pleasure (making laughter or

- teaching). . . . Too many in a group is such a strain on
 (8) the power to inhibit that such groups generally break up into pairs. [Implying the power to inhibit self-expression.]

The need for power is particularly transparent in the contribution of a male graduate student in psychology, an Australian with good but not brilliant intelligence, with a reputation for unlimited capacity for careful work and a somewhat naïve philosophical outlook derived from early religious experiences:

- I like to feel able to express the things in which I am inter-
 (9) ested am not a very good listener if the other person's interests do not fit in with mine. . . . Satisfyingness depends on adequate self expression.

A male psychologist in a mental hospital, about 30, very well poised, with excellent intelligence, a charming conversationalist himself, includes in his list of needs satisfied

- Self-expression of some sort—domination of situation at in-
 (10) tervals. . . . Satisfactory where participation general—no individual domination of situation.

A housewife in a middle western state, with a background of wise and progressive teaching and a reputation for originality, intelligence, and poise, believes in respect to motives that

- Actual motives are along the "me and mine" line; one would
 (11) probably wish to be less self-centered.

A woman of about 24, college graduate of high intelligence and calm personality, employed as a psychological secretary:

- The primary psychological need is the desire to impress one's
 (12) own personality upon the other there are three types of conversation, one in which A dominates, one in which B dominates, and a third in which A and B exchange equal honors.

An interesting sub-species of this motive is contributed by a male graduate student in psychology, age about 25, of brilliant intelligence and highly poised but somewhat reserved personality:

- Kinds of conversation that are particularly satisfactory;
 (1) intellectual, by which I mean not the display of sophistication, but the solution of abstract mental tangles. . . . It
 (13) seems that the display of knowledge and the verbal formulation of ideas per se should merit rank with the other types, but I think the difference is that insofar as this kind of thing lacks reciprocity, it is in that degree less conversational, even though it may be at times very gratifying.

It may be the same motive which another graduate student of about 35 has in mind; this man has a broad experience of life in a number of cultures, but his intelligence is probably not constituted in such a way as to yield a maximum of new ideas by the manipulation and recombination of abstract and intangible concepts; he was at the time of the inquiry much under the influence of the vocabulary of behaviorism, without much certainty as to the meanings conveyed:

- The language responses of other organisms bring forth either antagonistic, indifferent or approving responses on the part of
(14) the particular organism concerned. . . . when the language behavior calls forth a response on the part of the other organism involved that meets with the approving language behavior of the other, then it may be said that the approving organism is satisfied.

In the group of responses last mentioned, we seem definitely to have a real motive in conversation, viz., the ego-maximization consequent on socially approved performance. From the social point of view, however, the difficulty immediately arises that performance by one member of the group automatically prevents performance by others. The contributor of (13) above touches upon this, as also in the following:

- If in the lack of the conditions named it is still satisfactory,
(15) then it has very likely become less conversational and more in the nature of a monologue by the abler participant.

The psychoanalyst who contributed (8) above also seems to have the same difficulty in mind; he implies a convention, tacit but effectually enforced through the censorship of other egos awaiting expression, that the "floor" shall be passed around. One is reminded of the classical definition of an egotist as one who talks about himself when you want to talk about yourself. Needless to say, the maintenance of such a balance is delicate in the extreme; younger persons ordinarily lack the "feel" of the situation, and so take too much of the group's attention; a single individual animated by unconscious inferiority feelings for which he attempts compensation may ruin the situation hopelessly; a host or hostess, ideally a sort of "moderator" (excellent term!) entrusted with the smooth and equable distribution of the functions of self-expression, too often regards the whole situation as designed to provide an opportunity for public admiration of himself; and the degree of social anxiety has become so acute, at least in America, that it is the unusual social

situation which does not strongly resemble the conclave of a pack of wolves, each thirsting for his chance and restrained only by the ferocity of the others.

The individual opportunity can, of course, be thought of as the result of a simple division of the amount of time by the number in the group; accordingly, we find a very general impression that only in small groups is satisfying conversation possible:

"Small groups best; three or four ideal. With increase of group comes increased stiffness, restraint, etc."

"I like the small group or the lecture—but even this is better balanced by discussion time."

"Conversations are probably most satisfying when there are but two, three, or four persons participating, because there seems to be, except in exceptional cases, a tendency toward general incoherence concerning the subject matter."

"Conversations grow less satisfactory when the number of participants exceeds twelve."

"Ideal conversations in my experience are between two persons usually alone, sometimes with a third person present."

There are also two or three replies to the effect that if everything else could be equal, numbers would have no effect on satisfaction. In the light of the considerations mentioned, however, it is questionable whether any such condition is attainable.

The motive of interest in the other personality occurs occasionally; the following is from the contributor of (2) above:

The people taking part in a conversation are a more important factor than the topic. . . . Numbers do not affect the satisfactoriness of conversations as much as do the personalities of the participants.

From the contributor of (1) we have also

I do not believe that the nature of conversational topics is of any great importance except in so far as the topic may distract one's interest in the personality one is conversing with.

This interaction motive brings us rather naturally to the contributions of the three psychoanalysts included among the respondents, since it is only in their contributions that a definite attempt is found to make explicit the exact relationships between the personalities involved. Perhaps the least explicit and illuminating of the three comes from a man of about fifty, a student of Freud and something of a dogmatist; he is Jewish (European background), and has been in

practice as a psychiatrist for most of his professional life, both in institutions and privately:

- Adult conversation in the course of social intercourse is essentially an analytical situation and in such conversation the usual transference and resistance mechanisms enter. When a conversation is satisfactory, it may be said that there is a positive transference, when unsatisfactory, a negative transference (resistance) a conversation involving
- (18) personal matters is more likely to precipitate a resistance than one which is less personal, and furthermore, the person to whom the conversation is addressed may be a substitute in the mental life of the speaker for earlier personal imagery carried over from childhood. When a person is "bored" by a conversation an instinctive opposition is at work arising within the individual who is talking or listening. A "heart-to-heart talk" between friends is an example of good transference.

This is clearly a more penetrating point of view than we have been accustomed to find in the non-analysts; but it suffers from a deficiency often found in the users of analytical terminology, viz., the tendency to regard a phenomenon as explained when it has only been named or classified. In the first place, it is technically a confusion (though it is not certain here that the effects are profound) to identify a resistance with a negative transference; the former is an obstacle to the progress of the interchange, the latter implies hostility on some level. Negative transference may grow and bloom and fade without any marked resistance, although it is usual for it to be used, at least part of the time, as a resistance; resistance may be very stubborn without any marked negative transference. The substitution mentioned is the essence of transference; but it is far from certain that all satisfying conversation is simply positive transference, and the reverse; and in the cases in which it is so, the mechanism is almost as unclear as in the remarks of the non-analysts. For somewhat more light we may turn to the student of Rank mentioned above, the contributor of (8):

- The most useful distinctions I see are degree of ego development and erotic age. The person who is able to have the give and take relationship I mean by adult love will want the same kind of conversation—to avoid the extremes of being the silent receiver (baby) or giving endless streams of mental pabulum.
- (19) The boredom so often suffered comes chiefly from the difficulty of meeting persons who want to play our conver-

sational games by the rules we would set up; so it is pleasanter to limit one's talks considerably with persons with whom one cannot gratify the pleasure of using talk to express mutual love and common interest.

The transference element is plain here, as in the previous contribution; but there is here a glimpse of possible mechanism. In addition, the account of boredom is much more satisfactory, and falls in line with the much more superficial account in (1), instead of being the exact opposite, as is the preceding. The third analyst is also male, about 35, of brilliant intelligence but somewhat unstable personality; he has been highly successful both in institutional and private practice:

(2) Need for understanding intimacy (Mother?); that one's cherished notions should be found in the mind of another increases feeling of subjective certainty, and ministers to sense of security.

(3) Satisfactory in that they seem to convey more than the meaning of cold words, finally ending in silent agreement.

(20) (4) Unsatisfactory in that they are arguments, the differences in which have personal bases and are not admitted as such; or in that they deal with matters of no import.

(10) Yes; ideally one should be a good listener, giving his friend the occasion for self expression, catharsis, and bucking up of the ego, but actually I find I have a bloody logorrhoea.

The transference basis is also plain here, though not quite so explicit. The question is also brought to the fore as to what the relations may be between the transference interpretation and that of domination. It would be tempting to examine this topic in terms of the larger setting of Freud vs. Adler, but for the present it may be sufficient to suggest that transference appears to be the more fundamental phenomenon; an impulse to domination implies competition, and probably competition for something very badly needed or very vital—i.e., sustenance (largely symbolic on these levels). This brings us directly to the statement of (19) and the slightly more symbolized one of (20), where the sustenance is support. Such considerations lead immediately to a topic of tremendous social import; transference involves fixation, from which it is desired (impulsively rather than rationally) to be free; that is, it is essentially an intermediate condition between infancy and adulthood (from the standpoint of emotional development). The fact that a number of intelligent and experienced observers find the need to dominate

(with its implied transfer) at the basis of most actual conversations, including their own, bears witness to an appalling amount of infantilism at the basis of our society. This is a devastating, but not an incredible or even a new conclusion; it is admitted at one extreme in the most transparent manner conceivable by such products of the popular consciousness as the "Me and Mine" comic strip and the sardonic ruthlessness of Ring Lardner and the creator of the Bungles; and at the other by the researches in group analysis of Trigant Burrow and his students, who feel that practically the entire fabric of social intercourse has a fictitious basis, being governed by social images implanted in early childhood, and that the predominant activity being carried on is the constant attempt to complete early patterns—an activity which, since it is based on nothing more substantial than a fantasy, is doomed to constant disappointment.

These matters are so vital to all human experience that few more helpful studies could be imagined than one to discover or formulate what might be called the normal content of adult life. This field is exceptionally vague and ill-defined at present; the sole positive contribution with a bearing on this point in the present material is "the give and take relationship I mean by adult love" in (19). Burrow has said that the neurotic may often be regarded as the only normal person, in that he is the only one aware of a discrepancy between social image and social reality and uncomfortable about it; but such a remark, being not only somewhat paradoxical but almost entirely negative, can hardly serve as a definition. The issue is sometimes stated by the analysts in terms of an antithesis between narcissism and object love; but this leaves the criteria of object love *per se* uncertain. Lippmann in *A Preface to Morals*, Freud in *The Future of an Illusion*, possibly Holt in *The Freudian Wish*, and a few other writers, chiefly critics, have shed some light on the subject from the side of ethics; but the whole conative analysis underlying the ethics has not yet been approached. Could such an analysis be effected, it would probably be found that the results would be applicable without change to conversation, on the ground that the latter is to be regarded as merely symbolized conduct.

A large proportion of the respondents agreed in recognizing an inconsequential as well as a considered type of conversation, and most granted a legitimate function to the former:

"I find myself becoming more tolerant of what I formerly considered petty and inconsequential conversation, not because

I regard it as valuable in its subject matter particularly, but because I would recognize it as having some value as a sort of social lubricant, facilitating the extension of one's range of acquaintances, etc., and serving as a means of bridging over the gaps between more serious discussions in the conversations of individuals and groups."

"(5) Such as casual, or arranged?"

"Indeed there is a distinction between conversations; daily conversations are usually rather desultory, and without any distinct gain to anyone, but they do satisfy a human need for expression; they seldom arrive at any conclusion, but at their happiest they may supply a nucleus around which some ideal conversation may be built, although fortunately they are usually spontaneous and not engaged in from the standpoint of self-improvement."

"I think I have two reasons for talking; to entertain a group, or at least to offer suggestions on the basis of which someone else will do so, and to find out something which I desire to know (including here the discussion of a problem, the attempt to solve a difficulty, etc.)."

"There are three kinds of conversation that are particularly satisfactory: (1) intellectual . . . (2) facetious, involving an exchange of humor and witticisms, (3) familiar, which might be devoid of depth or brilliancy, but evokes a pleasurable affect because of a community of interests, knowledge or experiences."

"The following classifications of conversations occur to me: formal and informal, light and deep, serious and non-serious."

There are a few returns or sections of returns which do not fit readily into any scheme heretofore mentioned, but are of sufficient interest to be quoted; the following is from a university teacher of psychology, male, about 34, of great intelligence and originality and well-adjusted personality; he writes in the mood of formulating problems rather than answering the suggested questions definitely:

"What about the individual differences in persons in the degree to which they seek to make use of conversation? . . . I am aware of a very strong inertia against opening a conversation, . . . This avoidance mechanism is greatly strengthened by fatigue and by the presence of persons toward whom I have, on other grounds, a negative response. . . . In this way [described elsewhere] the threshold against opening a conversation is lowered, and for some time thereafter I am much readier to talk with anyone. . . . One other point should be mentioned—the vocal pitch which is 'naturally' employed on a given occa-

sion. . . . The amount of talking necessary in the day's work is more than sufficient to satisfy any social needs I may have. . . . Others as well as myself are unskilled in conversation as an art, and even if we were skilled we would be unable to get very far on the basis of common interests. Dinner party conversations are commonly forced, with evidence of showing-off, or of talking merely to fill uncomfortable gaps. It is a question whether the conversation isn't more embarrassing than the gaps. [Cf. the contributor of (10): Satisfactory where no necessity for 'Keeping up the conversation'—where periods of silence are the accepted thing.] The prevailing social patterns seem to favor superficial motives for conversation . . . we make little attempt to use conversation as a form of group thinking, or to exploit its analytic and self-revelatory possibilities. What are the inhibitions to conversation? I would like to see a study made of the development of conversational inhibitions in pseudo-family groups. What causes the gradual dying down of mutual stimulation? Monotony cannot fully account for it, nor the exhaustion of discussable ideas. In the field of conversation, among close friends and relatives, something seems to happen which is analogous to Hamilton's observations on the sex-attraction of monkeys."

The following is from the contributor of (7):

"The conversations I liked . . . were usually imbued with the spirit of 'We ought to get out and reform the world,' which I would now interpret as coming from a background of rather wide habits of self-condemnation and guilt. I counted them very satisfying conversations, though [since they relieved the guilt]. . . . When a person represents truly the distinctive flavor of his own life in his conversation—that is, not trying to force his peculiarities on others, but at the same time not trying to camouflage his opinions and past history under some conventional pattern—that is what I would say people are most likely to appreciate and rightly count worthwhile conversation. . . . I find fairly frequently that others are able to bring out aspects in the personalities of others that are interesting to me, but that I am myself unable to tap or lead around to, so that sometimes the conversations I enjoy most are those between others that I have the chance to 'sit in on.' Left to myself with the same persons I would sometimes come to sword's-points with them, or drift away from them in indifference because of difference in viewpoint on minor matters."

The outcomes of the study, while not definite in any sphere, may

be summarized briefly as follows: (1) three times as many men as women showed sufficient interest to reply, possibly indicating a greater need and ability on the part of men to rationalize their motives in this sphere (in the non-technical sense); (2) there appear to be several levels of penetration in thinking about the subject, from "The function of satisfactory conversation is to furnish satisfaction" to "Finding one's cherished ideas in the mind of another increases the sense of subjective security"; (3) there is a widespread feeling that one of the principal sources of satisfaction in conversation is the exercise of domination or display; (4) by the psychoanalysts responding (although of different schools and backgrounds) conversation is seen as a transference situation, involving at its least social levels a reproduction of the infantile mother-child situation and at its most social levels an adult give-and-take situation of the sort known technically as "object love"; (5) distinctions are widely recognized between light or playful conversation and serious or creative conversation; these might be thought of as analogous respectively to a dance, in which the satisfaction is derived from the kinaesthetic and other sensations, and creative work, where the satisfaction seems to reside in the successful objectivation of purpose on the external world; (6) suggestions by individual respondents not readily classifiable include (a) the investigation of individual differences in capacity to make use of conversation for expression, (b) the waning of motive for conversation with increased association, (c) conversation as a channel for the relief of guilt feelings, and (d) the necessity of fearlessness or confidence for the highest conversational satisfactions.

RELATED WORK

1. BURROW, N. T. The social basis of consciousness. New York: Harcourt, Brace, 1927. Pp. 274.
2. ELLIOT, H. S. The process of group thinking. New York: Asso. Press, 1928. Pp. 239.
3. JOHNSON, R. I. Determining standards in English composition. *School Rev.*, 1928, **36**, 757-767. (Particularly "Standards in conversation," 759 ff.)
4. SHEFFIELD, A. D. Training for group experience. New York: Inquiry, 1929. Pp. 119.
5. SOUTH, E. B. Some psychological aspects of committee work. *J. Appl. Psychol.*, 1927, **11**, 348-368, 437-464.
6. SYZ, H. C. Remarks on group analysis. *Amer. J. Psychiat.*, 1928, **8**, 141-148.

Clark University
Worcester, Massachusetts

LES FONCTIONS DE LA CONVERSATION

(Résumé)

On a envoyé à 25 hommes et à 25 femmes une enquête demandant des "associations libres" sur les idéals et les buts de la conversation des adultes; 15 hommes et 5 femmes y ont répondu. Parmi les réponses qui montrent une réflexion soignée, il y en a eu beaucoup qui appuient sur les satisfactions de l'importance du moi dans une bonne conversation; trois psychoanalystes décoles différentes et de milieux différents ont vu les situations de conversation comme essentiellement situations de transfert. Beaucoup des réponses ont distingué un type "léger" et un type "sérieux" de conversation. Parmi d'autres problèmes suggérés ont été ceux des différences individuelles dans l'élan vers l'expression dans la conversation, les causes de la décroissance de l'élan vers la conversation avec une association augmentée, et l'association du succès dans la conversation avec l'état d'être sans peur avec la confiance.

WILLOUGHBY

DIE FUNKTIONEN DER UNTERHALTUNG

(Referat)

Es wurde eine Anfrage an 25 Männer und 25 Frauen gerichtet und gebeten, "freie Assoziationen" über die Ideale und Ziele der Erwachsenenkonversation mitzuteilen. 15 Männer und 5 Frauen antworteten. Von den Antworten, die sorgfältiges Nachdenken bewiesen, hoben manche als wichtigen Faktoren in erfolgreicher Konversation hervor: die Befriedigung der Ego-Steigerung; drei Psychoanalytiker verschiedener Schulen und Erfahrungen betrachteten die Situationen der Konversation im Wesentlichen als Situationen der Übertragung. Manche Antwortenden stimmten in der Unterscheidung von "leichter" und "ernster" Art der Konversation überein. Unter anderen Problemen, die vorgeschlagen wurden befinden sich die nachfolgenden: die der individuellen Differenzen mit dem Trieb nach konversationellem Ausdruck, die Ursachen der Abnahme konversationeller Impulse mit zunehmender Assoziation, und der Zusammenhang des konversationellen Erfolges mit Furchtlosigkeit oder Vertrauen.

WILLOUGHBY

THE SCIENTIFIC STUDY OF HUMAN SEXUAL BEHAVIOR*¹

O. L. HARVEY

INTRODUCTION

Of that which today is being written concerning sex—by would-be prophets of a new era, as well as by equally blind antagonists of even the suggestion that men and women are anatomically different—much is irrelevant. Any emotional appeal, in either direction, must fall short of the ideal essential to the accurate perception of facts. For some pseudo-scientists “sex” constitutes an intense religious creed; to dogmatists of the other persuasion it is worse than anathema. It is possible, however, as recent publications have demonstrated, for human sexual behavior to constitute a subject for scientific investigation, in much the same way as has the study of the movements of stars or of the intestinal tract.

From the point of view of scientific impartiality the aim is not to discover whether “sex” makes people happier (whatever that may mean), nor whether it is to be deplored that such a distressingly embarrassing function as reproduction is essential to human survival, but rather to determine the pertinent facts in so far as they are discoverable, and, subsequently, to formulate laws concerning sexual behavior as they emerge from those facts.

It is indeed regrettable that so much unscientific jargon concerning sex reaches the public eye and the public ear. For an emotional appeal generally arouses an emotional response, whether of contemptuous disapproval or of abandoned enthusiasm. And, unfortunately, opinions formulated in a moment of emotional stress tend to persist even in the face of facts coldly adduced in qualification of fallaciously generalized concepts.

The need for a concise exposition and criticism of the different

*Recommended by Raymond R. Willoughby, accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office, April 29, 1931.

¹In the collection of these data, and in much of the subsequent treatment, the writer was ably assisted by Mr. Karl A. Fehr, at the time a graduate student in the department of psychology, University of Texas.

methods of investigation used by reputable scientists, judged in terms of their conformity to the simple yet inexorable requirements of logical thought, has led to the compilation of this short work.

The attempt has been made throughout to present the material from the point of view of orthodox methodology in a way understandable to relatively well-disciplined readers, who, though not directly acquainted with the most recent researches into human sexual behavior, nevertheless are interested therein, but find themselves at a loss to determine to what extent the "findings" of different authors may be accepted as fairly adequately substantiated by scientific proof.

PART I

THE PRINCIPLES INVOLVED IN SCIENTIFIC METHOD

From the point of view of applied science man seeks to know so that he may the better control or provide for the future. And progress in knowledge involves fundamentally the more accurate, more detailed, definition of facts, and the establishment of laws of relationship between them. The initial statement of a law is expressed as a hypothesis arrived at "intuitively" by some student. Thereafter it remains to test out the hypothesis by an examination of further facts; to attempt, on the basis of these facts, a precise prognostication of future events; and, accordingly, to confirm, amend, or reject the hypothesis advanced. Methodology in the realm of the biological sciences (with which the present work is concerned) postulates that, in so far as possible, certain fundamental requirements should be met; to the extent that any investigation falls short of them it lacks scientific value. These requirements may be briefly stated as follows:

1. *Sampling*

a. The sample group, with respect to which the investigation is conducted, should be as nearly as possible representative of the original population from which it was drawn; it should not, because of bias in selection, present a distorted picture of the facts.

b. The group which it is presumed to represent should be fairly homogeneous with respect to certain important variables (such as, within the limits of the present universe of discourse, race, sex, age, social status, and formal education).

c. The number of cases selected for investigation should be sufficient to warrant practical certainty that conclusions based upon

their examination will remain fundamentally unaltered with respect to any other similar sample of cases from the same population. Lacking this assurance, some generally recognized measure should be given of the degree to which findings derived from the sample in question may be assumed to be representative of those derived from the whole.

2. *Constancy.* In so far as possible the situation in which experimentation is conducted or observations are made should be essentially the same for every individual in the group. This requirement applies not only to the physical apparatus in, and the appearance of, the room in which the subject is placed, but also—and most strictly—to the presence, appearance, movement, and conversation of the experimenter, and to the medium of communication between experimenter and subject. A further factor, as yet unmeasurable, but nevertheless evidently of importance, is the emotional condition of the subject. The distorting effects of upsetting emotional states may be to some extent reduced by placing the subject at ease, in rapport with the examiner.

3. *Objectivity.* *Ceteris paribus* the less the recording of the subject's responses is left to the subjective interpretation of the observer or experimenter, and the more it inheres in the precise nature of some recording instrument, the better the method.

4. *Experimental Controls.* Of all methodological requirements, that which is the most important is the provision of adequate controls. It is not warrantable to assume that certain phenomena observed in the findings from a given experimental group are probably causally related to some "experimental variable," which has been deliberately introduced into the situation, unless it can equally be demonstrated that, in a group, fundamentally similar, into which, however, the same variable has *not* been introduced, similar changes cannot be observed. This process of establishing controls involves meticulous care, and is probably the most tedious operation in psychological experimentation, for in certain instances it involves an equalization of the groups (namely, the "experimental" and the "control" group) not only in general, but also with respect to every individual of the aggregate—a process known as "matching."

Where indirect observation alone is concerned, however, it is seldom possible to control conditions to this extent. Nevertheless, in such cases a precautionary control of a different nature is found necessary—a control which, because it inheres in the nature of the

sampling, is often neglected. A practical instance will serve to illustrate the point. A questionnaire is sent out to an impartially selected group, only a limited percentage of which, however, replies. Certain relationships are found to obtain between certain observed phenomena recorded in the replies received. Unless, however, the nature is known of the percentage of individuals who failed to reply, it is not warrantable to assume without qualification that findings based on the available data are applicable to the entire population which one originally canvassed, regardless of the fact that one's original canvass was impartially directed. The findings are applicable only within the limits of sampling, and need to be qualified accordingly.

5. *Statistical Treatment.* The three most important rules under this heading are as follows:

a. The degree of refinement of statistical techniques used in the treatment of data is necessarily determined by the degree of refinement inherent in the nature of the data themselves. The application of elaborate statistical devices to relatively crude observations is not only absurd but sometimes actually misleading, since it conveys an impression of accuracy of definition which the facts do not warrant. Similarly, the inadequate treatment of data, whether because of inexpertness or of lack of thoroughness, constitutes unsatisfactory scientific procedure. Always, where possible, complete basic data should be either published or filed away for the use of later investigators in some permanently accessible locality designated in the article or book containing the report of the investigation.

b. An instrument of observation, or measurement, should actually observe what it is claimed to observe. The more truly it does so—i.e. the more valid it is—the more satisfactory will it be as a scientific device. Validity is usually determined in relation to some external criterion, such as the judgment of experts in the field concerned. Established techniques already exist by which the degree of validity may be determined.

c. The more consistently does an instrument of measurement yield similar observations on the same general group of phenomena, as applied to similar samples of the same population, the more reliable it is in relation to other instruments of professedly like nature. Where possible an index of the reliability of the instrument should be given. Where not possible, that fact should be stated.

6. *Definitions.* Care should be taken that the more important

terms used should retain their definition throughout the investigation. Subtle changes in the meanings of terms used give rise to some of the most serious confusions in scientific thought.

7. *Findings.* Unqualified generalization, from findings derived from limited data, or specific to a limited universe of discourse, constitutes one of the most serious crimes against scientific procedures. Reasoning from analogy or from the particular to the general may be subsumed under this heading.

Have the works of the foremost investigators into human sexual behavior complied with all of the above requirements? No, they have not. To conduct an investigation in which all of these postulates were observed would constitute a tremendous task. The individual investigator attempts to ensure compliance with as many as possible or as seems to him to be sufficient for his immediate purposes. The remainder he introduces, or should introduce, as necessary qualifications on his findings. Not all investigators have done so. In many cases the qualifications have been provided by critics. In general, however, it is in this way, by gradual accretion and continuous readjustment, that complete understanding is approached.

In the following chapters an attempt will be made to indicate the extent to which different techniques of investigation, as represented by the work of certain outstanding investigators, conform to the above requirements of orthodox scientific method. In particular, deficiencies will be indicated. It should be clearly recognized, however, that criticism of this sort is directed at no particular method or individual. The aim is purely technical and general. The point of reference is one of objective scientific requirements, devoid of personal preference. Essentially the aim is *not* to deprecate the work of others, but rather, in pointing to the deficiencies in their technique, to make critical judgment possible, and to suggest to future investigators the pitfalls to avoid.

PART II

THE METHOD OF DIRECT OBSERVATION

The study of the sexual behavior of human beings follows two lines of scientific procedure: direct observation, and personal report, the differentiating factor being the use of language in the latter.¹

¹It is evident that any investigator might use either or both methods. The following classifications are in consequence to that extent artificial. It is to be understood, however, that the inclusion of any investigator's name within one category does not necessarily exclude it from the others. The examples selected are intended purely for illustration.

In essence the method of direct observation involves an observer,² who reports on the behavior of an organism operating under more or less controlled conditions. Usually also it involves the observer's interpretation of his own observations. From the point of view of scientific precision there are four different kinds of direct observation. Each is briefly described below in terms of its principal characteristics. Obviously, the degree to which these characteristics apply to any specific study will vary slightly from one investigation to another; the descriptions here given attempt to provide a composite sketch, an outline, of the best and most recently developed practices, not a detailed description of every feature. Under each heading, however, an attempt will be made, where necessary, to point to these minor, yet nevertheless significant, differences.

1. *The Synthetic Procedure*

Large selected sampling;³ many different observers; situations seldom the same; subjective interpretation of observations by observers often untrained; no control group; no statistical treatment of results.

Under this heading may be found the works of such writers as Briffault (4), Ellis (10) Moll (36), and Westermarck (43).

We have here, as it were, a synthetic portrait of human behavior, the resultant from a succession of similar refinements, the original material being the reports of travellers and others, who claim to have actually witnessed certain of the forms of behavior which they report. The procedure of these eye-witnesses is comparable to that recorded in the next following section, excepting, however, that the latter observers are scientifically trained. The essential difference between these two procedures is primarily one of evidence. Other things being equal, it seems feasible to suppose that the synthesis of the hearsay evidence of many untrained witnesses is not likely to be so precise as the report of a single trained eye-witness. A further difference is to be noted in the extent of the field covered by the respective methods of observation. Thus, by comparing the reports

²In view of the fact that an experimenter is one who observes phenomena under controlled conditions, the term "observer" as used here should be taken also to include "experimenter."

³To mark the difference between on the one hand, a sampling which may be considered appropriately representative of, and chosen without bias from, the population from which it is drawn, and, on the other hand, a sampling either unrepresentative, or selected as the outcome of bias, the terms "representative" and "selected" respectively will be used throughout this work to designate the two kinds.

of witnesses of the same sample, the "synthesist" compensates to a considerable degree for the defects inherent in his procedure; and, at the same time, tends to check any bias which, because of his observations made on a limited homogeneous sampling, the worker in a field of more limited scope might manifest.

It is obvious that control groups and statistical treatment of findings are not appropriate to this procedure.

2. *The Anthropological Procedure*

Adequate sampling, selected from highly homogeneous population; observer and situation fairly constant; subjective interpretation of his own field-notes by a trained observer; no control group; no statistics.

Here are to be found studies by Malinowski (32) and Mead (34).

Although field investigators are concerned primarily in the observation of group behavior, they trace its operation necessarily through the medium of the behavior of individuals. They are, however, also able to report occasionally on group behavior, as, for example, in the dance, where certain courtship phases of sexual behavior may be studied. The observations of both anthropologists here mentioned have been confined to small limited village communities, fairly effectively isolated in so far as the influence of disruptive foreign mores is concerned. The situation is thus held constant within fairly close limits. It is to be noted, however, that in this type of investigation the observer possibly constitutes a significant factor in the total situation, and should be evaluated as such by the observer himself. The validity of an observer's findings will depend to a considerable extent upon the recency of his note-taking after the event, and the degree to which the notes were objectively worded. It is important that the actual behavior of the subject, rather than the motives or drives which, to the observer, may seem to impel him, should be recorded. And, when later it comes to the interpretation of his notes, it is even more important that personal "unconscious" bias should not color his findings. The presence and behavior of the observer under the circumstances which he is reporting should be taken into account where possible. As is pointed out in the next section, this important factor can be too easily overlooked.

The observations made by parents, teachers, and nurses concerning the sexual behavior of children under their charge may sometimes, with appropriate qualification, be included under this category.

3. *The Comparative and Child-Study Procedure.*

Inadequate sampling; observer constant, and situation thoroughly con-

trolled; interpretation of own notes by experienced observer; no control groups; no statistical treatment.

Under this heading are to be found the works of Bingham (3) and Hamilton (16).

Under the circumstances of this procedure, the subject is placed in a prepared compartment, to which it is accustomed, where its behavior is noted by an observer, who at will becomes either a part of the situation or a concealed observer outside of the observation chamber. Other subjects, of the same or of the opposite sex are introduced at will by the observer, who then records the behavior of the subjects in response to the altered, but controlled, situation. By controlling at least to a considerable extent, the sexual activities and stimuli of the subject and his companions, the observer is enabled to make his interpretations the more precise. In this, as in the preceding section, there is the danger that, unless adequate precautions are taken, the observer will become a significant and uncontrolled variable in the total situation observed.

It is to be noted that neither of the studies here quoted deals with human beings. In pediatric and educational literature sporadic reports are to be found relating to the sexual behavior of children and infants. None of them, however, so far as the writer can discover, is as carefully controlled as the studies here quoted. Hamilton's study deals with monkeys and baboons; Bingham's with chimpanzees. Nevertheless, the methods there outlined are of such significance as to warrant careful study by observers of human sexual behavior. In this connection Bingham writes, however, "Man has unexcelled facility in sharing and concealing his own mental content. Direct study of his spontaneous sexual behavior during childhood becomes more and more difficult as his mentality increases. During early adolescence such study probably becomes quite impossible" (p. 158). There is, nevertheless, the possibility that, by some adaptation of the procedure of child-observation used by Gesell at Yale, at least the pre-adolescent sexual behavior of children could be successfully studied. A similar procedure applied to older individuals would probably incur for the would-be observer intense social disapprobation—but so, in the early history of medicine, did the activities of Vesalius and his immediate successors, who, to obtain material for their profoundly significant researches in anatomy, are reported to have gone to the extent of robbing graves and gallows of their human relics.

The nursery school and the home provide well-controlled, but relatively innocent and fertile fields for fairly precise observation of the sexual behavior of young children, a procedure comparable to some extent to the observational method used by Hamilton with respect to monkeys, which were allowed to range fairly freely within a large enclosure, where, however, they were ever within sufficiently close range of the observer.

The number of cases observable under the conditions of this method of observation do not permit of statistical treatment.

4. *The Laboratory Procedure*

Fairly adequate sampling; observer constant, and situation thoroughly controlled; objective interpretation relative to normative standards; controls often contained in the experimental procedure itself; statistical treatment varies with the nature of the study.

The works of King (25), Lipschutz (29), Marshall (33), Miles and Terman (35), Pressey (41), and Kent and Rosanoff (24) subsume under this category.

Although superficially it would seem that studies of the internal secretions of the sex glands are in no way related to those concerned in sex differences as determined from tests conducted in psychological laboratories, nevertheless, in their methodology they are in essence closely similar. In each case the observer (considered as "surgeon" or as "tester") is entirely subordinate to the experimental set-up itself. Whether from photograph and anthropometric data, or from scores on tests and verbal responses to simple verbal stimuli from standardized lists, the interpretation of findings is fundamentally objective, determinable on the basis of a normative pattern or scale. Even incidental observations, additional to these, are fairly objective (e.g., in test situations, fidgetiness, slowness in response, and negativism; in surgical situations, the growth of hair on the face or pubis, localized depositions of fat on the torso, etc.).

It is obvious that, in view of the less risk, expense and expertness required in psychometrics as distinct from surgery, the larger number of cases will accrue to the former. Comprehensive data on the latter are not provided. Here we have the most precise form of direct observation. The situation is strictly controlled, the observer himself being replaced by recording devices; experimental variables are introduced objectively; and the interpretation of findings is limited to the comparison of the behavior of individuals relative to an established scale of norms, with respect to behavior patterns of limited range.

Thus far, then, what have we? The "synthetic" procedure of direct observation appears to be the one most open to criticism. It amounts to little more than the suggestive but scarcely reliable statements of observers not trained in the objective methods of investigation into sexual behavior. In the "anthropological" procedure we have trained observers contributing exceedingly valuable observations concerning group behavior, but necessarily lacking the objectivity of inquiry, the interpretation of records, and the accumulation of comparable data applicable in the same way to every member of the group, that are essential to scientific investigation. The "comparative" procedure promises exceedingly well, excepting for the severe limitations of numbers. Furthermore, it applies, as yet, to sub-human primates only. Direct observation of human sexual behavior is satisfactorily possible only during that stage of life in which overt sexual behavior of any direct social significance is at its minimum. And, finally, although in the "laboratory" procedure we discover the most precise of direct observation techniques, it appears either to be limited to relatively small numbers of cases, and to apply to physiological studies far divorced, in the phenomena with which they deal, from everyday sexual life, or to be limited to observations of incidental behavior in a situation which is essentially one of personal report.

Thus, excepting in the field of anthropological research, it would appear that direct observation is not likely to contribute much to our stock of knowledge concerning human sexual behavior *per se*. Child study seems to offer the most promising prospects out of the remaining fields. A technique of direct observation sufficiently comprehensive, yet still consistent with the requirements of scientific method, has not yet, however, been developed. Possibly nursery schools and the homes of intelligent parents will provide the first suggestive clues.

PART III

THE METHOD OF INDIRECT OBSERVATION OF PERSONAL REPORT

Complementary to the method of direct observation is that of personal report. Here observation is limited to recording the relatively uncontrolled verbal expression of the subject; it is concerned with language behavior, rather than with gross bodily behavior. Significant also is the fact that the stimuli provided by the observer are also expressed in language. The procedure comprises not merely

the recording of the response, but also of the verbal stimulus which elicits it, both being expressed in language. Furthermore, whereas in the method of direct observation it is the present alone with which the observer is concerned, in personal report it is necessary to note the behavior of the subject unlimited in the flexibility with which he may extend himself in time. A serious obstacle in the method of personal report lies, however, in the fact that recall memory is not always certain. Many questions involve opinion rather than fact; and it is notoriously the case that subjects tend to conceal rather than reveal their socially less-approved thoughts. Thus it is evident that in the method of personal report an observer is on far less certain ground than when directly observing behavior. On the other hand, exigencies of human living do not permit of lifetime studies of individuals; whereas many years may be condensed into a single sentence. The method of personal report is exceedingly flexible. Furthermore, much that cannot be observed directly can, nevertheless, be reported on, presumably with considerable reliability, by willing and honest subjects. Consequently, the liability of error due to faulty memory or the distortions of rationalization would appear to be insufficient to discount the method entirely.

In the same way as in the preceding part, the procedures included under this method are classified below.

1. *The "Anecdotal" or Autobiography Procedure*

Large unrepresentative sampling; trained observer, but many different situations; subjective interpretation by observer; medium of communication uncontrolled; no control group; no adequate statistical treatment of results.

In this classification may be placed Ellis (10), Krafft-Ebing (26), Lindsey (28), Lipschutz (29), Malinowski (32), Marshall (33), Mead (34), and Moll (36).

The procedure here is essentially that of general discussion: observer and subject exchange remarks, the nature of which is determined by the exigencies of the moment. The function of the observer is usually that of asking questions; that of the subject is to answer them. The more skilled the observer, and the more confidence placed in him by the subject, the more likely is the observer to obtain comprehensive, detailed, and trustworthy data. It is evident, however, that the relationship (*rapprochement*) between observer and subject is vital to the entire situation. It is relatively seldom that the observer obtains a written statement from the subject which is as

complete and frank as an oral utterance; and in the latter case it is seldom that a verbatim report of the conversation is made. Consequently, the interplay of personalities, especially the part played by the observer in molding the subject's responses is, generally speaking, uncontrolled.

Although it is possible in this procedure to classify roughly to some extent the data which accrue, in the individual observations thus far made practically no statistical treatment of results has been effected.

2. *The Psychoanalytic Procedure*

Size of sample uncertain; selected population; fairly constant situation; observer variable; subjective interpretation from unreliable instrument; no statistical data reported; no control groups.

Here are to be found such investigators as Adler (2), Ferenczi (12), Freud (13), Jung (23), and Stekel (41a). Healy (20) provides an excellent summary of this technique. Knight Dunlap (8) criticizes it from the viewpoint of scientific psychology.

It seems fairly certain that the large majority of subjects, under psychoanalytic treatment, is drawn from the wealthier strata of the population. That the sampling is probably still further selectively determined is evidenced by the fact that psychoanalytic treatment is usually sought and continued by the subject voluntarily. That this is due not to economic causes alone seems fairly certain; it is possible, on the other hand, that psychoanalytic subjects are temperamentally peculiarly liable to be attracted to psychoanalytic procedures. It is suggestive in this connection that some subjects so early discontinue treatment. It is unfortunate that control group observation has not as yet been found feasible.

The term "sexual," as used by psychoanalysts, is liable to various interpretations.⁴ Sometimes it is used in a narrow sense, being limited to copulatory play, or still further to sexual intercourse itself; at others it is used to connote also activities as distantly related to these as are alimentation and elimination. For this reason it is exceedingly difficult to follow precisely the significance of psychoanalytic findings. The value of psychoanalysis as a scientific (i.e., explanatory, and ultimately prognostic) method is to this extent limited.

It appears that the environmental conditions of the psychoanalytic

⁴For an illustration of this point see (45). The writer hopes, in some later article, to offer a tentative definition based on arbitrary, but nevertheless comprehensive, behavioristic concepts.

clinic remain, as regards the individual psychoanalyst, fairly constant. Usually the subject reclines at ease on a couch in a room in which distracting stimuli such as the lighting of the room and the presence of objects within the subject's range of vision have been reduced to a minimum. The subject is directed to give his thought free rein, and to tell just what comes into his mind, without attempting in any way to criticize or direct his mental processes. At such time as tests are given (e.g., standardized word-association tests) the immediate setting is made even more constant.

The observer, on the other hand, would seem to be the most significant factor determining the outcome of the treatment. He directs the trend of the subject's thought along lines which appear to him, as a result of his experience as a clinician, to be the most likely to result in exposing "hidden complexes." As Healy and his co-authors express it:

"A certain amount of interpretation or explanation of the material furnished by the patient is supposed to be offered at strategic points by the physician. . . . Freud points out the necessity for the analyst to communicate to the patient the connections he expects him to discover but says it must be done at the right place E. Glover admits that one of the greatest difficulties in analysis is when to speak and when to be silent, and how much and how little to say, Ferenczi acknowledges that interpretation deflects the course of free association, both by rousing the ideas expected by the analyst and imposing association-prohibitions. However, he points out that the interpretations offered are not irrefutable; their validity depends on whether they can be verified by memory material, or through repetition of earlier situations. Also, he says, the analyst always remains somewhat skeptical about his own interpretations, and 'must ever be ready to modify them or withdraw them completely' " (pp. 432-433).

Concerning this same significant point, Hamilton (1929) writes as follows:

"The procedure employed by the Freudian analysts is as follows: . . . (d) At some time after the session the analyst records from memory the patient's verbal productions and any behavior which may have impressed him. (e) From time to time the analyst breaks his own silence by interposing interpretations of his patient's free association in terms of what these imply as to the nature and origin of the underlying adjustive

impulsions of which they are expressions in overt consciousness. A good analyst will often sit through many sessions without saying anything but in the end many such interpolations occur in all cases, and the free associations which the patient produces *after his preceding ones have been interpreted for him by his analyst* are regarded as scientifically valid material from which to infer the nature of his underlying adjustive processes. (f) The inferences which the analyst makes from his patient's verbal productions are either the former's common-sense reactions to them or direct 'intuitions' which come from the analyst's own 'unconscious' into his 'conscious' as inspirations of the moment. I wish to stress the point that the inferences which are currently arrived at in this lawless manner are currently exposed to the patient whose resistance to accepting them as true of him are discussed from time to time throughout the course of his analysis" (p. 14).

This might be effective clinical practice; it is in no way controlled scientific procedure. As Hamilton again expresses it, "the free associations obtained from a patient by such methods can have no scientific validity for purposes of comparison with those obtained from other patients" (p. 15). Nor, equally, can hypotheses developed on the basis of findings, derived in this partial manner, be looked upon as an adequate scientific explanation of the phenomena.

Fundamental to psychoanalysis is the technique of symbol-interpretation. Indeed this practice would appear to constitute one of the most powerful instruments of the psychoanalytic procedure. The interpretation of phenomena as they occur in practice is, however, subjective (with the exception of the limited interpretation permitted to responses to standardized word-association techniques). Especially is this to be noted in the common psychoanalytic practice of dream-interpretation. To induce the recall of experiences long "forgotten," it has been found time-saving to resort to a study of the subject's dream life. The concepts therein contained provide foci of sentiments ("complexes"), whence arise trains of associated ideas. These, if uncritically followed by the subject, reveal experiences, presumably forgotten. "The dreamer is instructed," writes Healy, "to talk about details of his dream, and specific trends and themes will emerge about which the associations will center; these lead finally to the discovery of the latent meaning in the unconscious content" (p. 432). Clinical experience confirms the observation that certain concepts commonly found in dreams—whether they be as objects or

as forms of behavior—tend, in a surprisingly consistent way, to initiate trains of associated ideas, which, proceeding along almost stereotyped lines, arrive ultimately at certain kinds of experience which are frequently sexual and are closely similar to each other in nature. So consistent, presumably, has been this observation, that clinicians have come to look upon it as warrantable to suspect certain dream-concepts as having symbolic significance; and, consequently, to pick them out for special investigation, directing the subject's attention particularly to whatever ideas they might find freely associated with them. Frequency tables substantiating this point, and similar to those published with respect to the Kent-Rosanoff word-association test have not, unfortunately, been published. Consequently, the validity of the technique as a scientific instrument can be accepted for the present as of doubtful validity only. The psychoanalyst does not, incidentally, adhere to a rigid table of symbol-interpretation, a fact which reduces its value as a reliable scientific instrument.

It would be of considerable value to have in published form the complete detail of a chance representative sample of psychoanalytic cases, in which a verbatim report of everything said by both subject and observer throughout the treatment was reported. Only in this way could an objective study of the authenticity of psychoanalytic procedures be effected. On the basis of such evidence it could be determined to what extent the observer does or does not determine the nature of the outcome. The partial report, which is customarily given, savors of special pleading.

No statistics of successful cures as contrasted with failures has been published by psychoanalysts.⁵ Uncertain as such data might be, because of the almost insuperable difficulty and expense of a consistent follow-up scheme, they would, nevertheless, constitute valuable information. An excellent example of the manner in which this could be effected is available in the clinical reports of E. Jacobson, published in his recent work on *Progressive Relaxation*. It is possible furthermore, that some attempt could be made to classify cases in terms of (for example) age, sex, number of hours of psychoanalytic treatment, techniques used in analysis, and (in a general way) the types of personality, or varieties of mental "illness" concerned. It is difficult, if not impossible, to discuss the qualities of a method unless one has relevant objective data to which reference may be made.

⁵See, however, the following recent publications: (18a) (42a) (44).

3. *The Case-Study Procedure*

Fairly large sample, representative of a homogeneous but selected group; situation fairly constant; trained observer; subjective interpretation of responses; measuring instruments used occasionally; medium of communication variable; careful investigation of subject's background; no control groups; fairly comprehensive statistics, but of no indicated reliability.

Under this heading subsume case-studies reported by Healy (19), Krafft-Ebing (26), and Moll (36) and by other clinicians such as Thom (42), Lindsey (28), etc. Under the same heading fall in part clinical reports (published and unpublished) by Dickinson (7, 7a).

The material which deals with human sexual behavior, under the heading of the sociological case-study procedure, is limited mainly to the work done by clinics, such as the Judge Baker Foundation and the Psychopathic Hospital in Boston, and the Institute for Juvenile Research in Chicago. Not every case is brought to the notice of the clinic because of sexual misbehavior, but in most cases the sex life of the individual is investigated as part of the regular procedure. Thus the publication of findings, from these clinics, deals with sexual behavior only incidentally. Nevertheless, the procedures here used in the investigation of phenomena, seeking not only to remedy existing defects, but also ultimately to determine the principles underlying behavior, are of interest in that as procedures they differ from those heretofore described.

The thorough sociological case-study involves an investigation into every available aspect of the subject's heredity and environmental history. Every type of procedure is used, from the "anecdotal" interview to the more strictly laboratory procedure of mental testing or of administering a standardized word-association test. Nor is the subject alone examined; a body of trained social workers investigates the subject's social, economic, domestic, and scholastic background. Consequently, at the final summing-up of the case every available detail of information concerning the subject has weight in determining the final diagnosis. Sexual behavior is observed in its proper perspective.

The procedure used in medical clinics by Dickinson and his co-workers involves not only careful measurements of the human genitalia in various conditions; but also the accumulation and analysis of many hundreds of reports, made as objectively as possible by scientifically minded subjects, concerning their sex history and behavior. On the basis of such data Dickinson has recently published material dealing with the sex life of women.

The number of cases reported in case-study investigations is fairly large; but as a sampling it is probably selected, depending upon the type of clinic concerned.

The sociological "case" is seldom, in the beginning, a voluntary subject; the ability of the investigators to make the subject feel at ease while under observation becomes, in consequence, quite a significant feature in the procedure. "When the interview has been secured," writes Lundberg, "its usefulness will depend largely on the personality and skill of the interviewer" (30, p. 153). So far as the interview phase of the procedure is concerned, the situation is likely to remain fairly constant. Where strange, or cooperating, observers are brought in on the case, however, a variable factor is introduced. In a similar manner, since different techniques are used in the interview, depending upon the nature of information sought by the interviewer, the situation will vary from subject to subject. So far as the field-work is concerned, it is evident that constancy of situation is impossible.

Only in so far as objective instruments are used in various phases of the interview does the observer become a constant factor. Generally speaking, however, no special effort is made to control this factor other than that exercised by the observer himself. Dickinson, being more particularly interested in precise physical measurement, uses measuring instruments to a considerable degree. The information provided by his subjects, on the other hand, is necessarily variable.

Brief reports of interviews are made by each worker concerned in the case-study investigation. The more highly trained the interviewer, whether clinician or field-worker, the more objective the report is likely to be. Nevertheless, it is evident that an unmeasurable degree of subjective interpretation is inevitably introduced into the report, which in its turn, is liable to subjective interpretation by other persons. "All the subtle influences of personal suggestion," writes Lundberg, "now become operative" (30, p. 153). Only the data derived from actual measurement—e.g., psychological test-scores, the number of persons resident in a given house, the timed duration of intromission during copulation, etc.—may be deemed adequately objective.

Comparisons have been made in a general way (e.g., by reference to findings from other studies) between subjects having similar histories but not all becoming sexual problems, and of subjects becoming sexual problems but having different histories. No report of an

extensive study has as yet, however, been published in which control groups have been used deliberately.

"The chief obstacle to the scientific utilization of case records and life history documents," writes Lundberg, "is the difficulty of treating them quantitatively—generalizing them statistically. . . . Case studies become of significance scientifically only when they are classified or *summarized* in some way so that the *uniformities* in large numbers begin to stand out and group themselves into general patterns or types" (30, pp. 173, 175). Yet the fact remains that few case studies are developed in conformity with a constant pattern. That a uniform procedure can be successfully adopted has been demonstrated in other fields of sociology. And leaders of thought in the social sciences undoubtedly are guiding activity in that direction. In the field of sex behavior, however, but for Dickinson and Healy, who have published the most comprehensive data, there is very little that is statistically of value.⁶

The publication of original data under the case-study procedure, has been necessarily limited to summaries, even to the summary of summaries, for each individual, with perhaps a special treatment of that phase of the study which most significantly illustrates the nature of the particular problem. The fact that a complete objective record of the investigation is not made, to that extent reduces the value of the case-study procedure from the viewpoint of scientific method.

The very comprehensiveness and thoroughness of the case-study procedure insure a considerable degree of validity. In this respect the case study is probably superior to any other of the procedures included under the major heading of "personal report." Whereas others confine their attention primarily and sometimes even exclusively to the sexual life of the subject, seeking other data only as they seem to be relevant to this particular phase of human behavior, the case-study procedure sees sex behavior as only a part of the whole, and thus is less liable to result in distorted findings and hypotheses.

The reliability of case-study procedures is determinable probably only in so far as other clinics might at some other date report on a case diagnostic findings similar to those reported by the first. For this reason any measure of reliability is difficult to obtain. It is

⁶As an excellent illustration of statistical handling of case-study material, in which a serious attempt has been made to determine validity and reliability of data, see a recent study by Ackerson (1a).

possible, nevertheless, that on a small group of cases such information may be actually available.

4. *The Questionnaire Procedure*

Large sample; selected population; situation and observer constant; interpretation from objective and constant instruments; statistical treatment fairly thorough, but of unknown reliability; control group possible, but not yet reported.

Under this heading are to be found Davis (6), Exner (11), Hall (15), Hamilton (17), Pearl (37), Peck and Wells (38, 39), Popenoe (40), and Chassell (5).⁹

The special advantage of the questionnaire procedure is that it elicits a considerable amount of information from a large population in a short period of time. Its chief defect, however, inheres, paradoxically enough, in just this same advantage; the information received is likely to be superficial. In the usual clinical situation it is possible to cross-question a subject subtly in such a way as either to confirm the accuracy of, or to obtain the qualifications which are appropriate to, his original statement. With the questionnaire it is sometimes possible to insert a few items constituting appropriate checks on other items in the same questionnaire; or, at times, to obtain a check on the replies elicited on one questionnaire by asking for replies to another, submitted at some later date, in which, although the information sought is essentially the same, the language in which the questions are couched is different. Questionnaires, however, have to be as short as the investigation will permit. The unwillingness of subjects in general to respond to a long questionnaire makes every item precious; with the result that one is tempted to forsake thoroughness in favor of comprehensiveness. The danger of superficiality can to some extent be mitigated (*a*) by guaranteeing to the subject that the questionnaire is entirely anonymous; (*b*) by establishing in the subject a favorable "set" concerning the investigation as a whole (this can sometimes be effected by a personally addressed, confidential statement, enlisting the frank cooperation of the subject); and (*c*) by wording every item in the questionnaire not only with a view to its simplicity, but also to the tactfulness with which it is expressed.

The questionnaire, being a printed instrument, provides a con-

⁹For a more extensive treatment of the questionnaire procedure see an article by the writer in the *Journal of Abnormal Psychology*, March, 1932.

stant, objective stimulus. Its objectivity will obviously depend to some extent upon the wording of the questions. It is, however, possible to make these "fool-proof" to a degree partly determined by the nature of the response sought—whether involving recent or remote recall, or recognition, an expression of opinion, or, as earlier indicated, an intimate confession involving an embarrassing self-revelation. The constancy of the situation, on the other hand, depends upon the manner in which the questionnaire is presented to the subjects. Davis, for example, mailed her questionnaire; consequently, her subjects responded each one (presumably) in a different immediate environment. Peck and Wells, on the other hand—primarily because of convenience in administration, but also because the nature of the entire investigation made it possible—presented their questionnaire to their subjects collectively and simultaneously in the one hall. Hamilton, seeking to avoid the impersonality of a mailed questionnaire, and to establish that degree of rapport which, on the other hand, is essential in the nature of clinical procedure, submitted his questionnaire personally to each of his subjects separately, but to every one in the same consulting room, in which even the distance between the chairs of the subject and the observer was held constant. To what extent any one of these three procedures yields the most complete and accurate responses is a matter which has not been experimentally determined. *Prima facie* it would seem that Hamilton's provides the most constant situation, combined with the highest degree of personal rapport. It involves, however, the longest time for administration, and considerable expense.

The observer, in the questionnaire procedure, becomes a constant and practically insignificant factor in the observational situation. In Hamilton's study this was insured by a brief initial conversation with the subject, in which incidentally the need for constancy of procedure was indicated. The individual items were then presented printed on cards, the subject's response being recorded by the observer who remained sympathetically mute throughout the interview, commenting only in so far as necessary to explain the meaning of words used in the questionnaire.

The behavior of the subject, in response to the questionnaire, is also objectively recorded. Generally speaking, the record constitutes a written response, entered by the subject himself. Hamilton's procedure differed from this merely in that here the observer made complete written notes, often verbatim, of the subject's responses, which

were given orally. This procedure has the advantage of releasing in the subject a flow of thought unhampered by slowness and the "inhibitions" commonly experienced by a writer as contrasted with a speaker. It has the possible disadvantage, however, that even a trained observer is likely, unless making a complete stenographic record, to report only those statements which appear to him to be significant.

The important problem raised in the last sentence concerns the interpretation of the subject's behavior—that is, of his verbal response. The customary type of questionnaire provides, after every question, a blank space in which the subject enters his response. Depending on the nature of the question, the degree of objectivity of this response will vary from a check-mark or a simple record of fact to a freely expressed, ambiguously worded statement of opinion, or a vague description of emotional disturbances. Peck and Wells attempt to control ambiguity by arranging their questioning in the form of a multiple response (including "yes-no") list of questions, requesting subjects to check the answer appropriate to their own particular case. There are, however, three objections to this procedure: (a) Some subjects, it is claimed (though unsupported by experimental proof) find it too impersonal. Their responses lack deliberateness and intimacy. The multiple-choice arrangement is looked upon as a time-saving device rather than as a means to facilitate objective classification. This defect can be countered to some extent by the degree of rapport initially established between observer and subject. Furthermore, it is obvious that the character of the experimental population will determine to some extent the veracity of the response. Seriously-minded intelligent subjects are likely to treat any form of scientific questionnaire with due deliberation. (b) Another objection directed against the multiple-choice modification is that a limited classification of possible answers militates against flexibility. This objection is met by including "any other" answer as an alternative. There remains on the part of the subject, nevertheless, a tendency to yield to the easy temptation of the alternatives offered, and to check one or the other item without qualification. (c) Finally, there is perhaps a tendency for the multiple-choice type of questionnaire to suggest to the subject a response which, had recall memory alone and not recognition been involved, he would otherwise not have considered appropriate. The seriousness of this defect—

assuming the objection to be pertinent—will depend upon the nature of the question.

A recent study by Chassell, dealing primarily with the causes of personality differences, but, incidentally, to a considerable extent also with sex life, constitutes probably the mechanically most nearly perfect sex-questionnaire yet published. Not only is there provided a choice of answers, but also provision is made for variations in behavior incidental to growth—e.g., behavior during childhood as distinct from that during adolescence. Questions are so arranged that no ambiguous answer is possible. This procedure is, in some respects, the antithesis of Hamilton's, and is liable to all of the objections referred to above. Its value as a time-saving device and statistically effective procedure, however, cannot be questioned. As a useful compromise between Hamilton's procedure and that of Chassell, it is recommended that, at some time prior to interview, a check-list, constructed on the lines of Chassell's questionnaire, should be given to the subject to complete. The more embarrassing questions, or those concerning which it is desirable not to suggest possible answers, or in which, again, it is desired to have the subject respond at length in an uncontrolled situation, could then be administered according to Hamilton's plan.

The questionnaire procedure facilitates the classification of data and statistical analysis. The multiple-choice procedure lends itself more favorably to this treatment than does that, as an example of the opposite extreme, preferred by Hamilton, who published long tables of responses with hardly any patent classification whatever. Probably a valuable and trustworthy mean could be established by developing carefully considered, strictly multiple-choice questions constructed from extensive detail such as that published by Hamilton. The tracing of "antecedence-subsequence" relationships from original data, so far as the published studies here considered are concerned, has not been entirely satisfactory. Wells reports only vaguely concerning a few correlations established between phenomena investigated by him. Hamilton supplies a few highly suggestive correlations, and tables showing percentage differences, and reports that, for thorough statistical analysis, he is at present engaged in tracing "patterns of behavior" by a process of sub-tabulation. The procedure promises to be of value. The number of his cases, however, is so small that the outcome of his investigation is likely to result (as he himself indi-

cates) in suggestion only, and not in convincing evidence. Other than Chassell's work, which is not primarily a study of sexual behavior, Davis' offers the best statistical treatment of results so far published. Unfortunately, however, this excellence is obscured by the somewhat complicated manner in which the material is arranged. To make possible clear and speedy comprehension of the significance of her data a considerably more simple arrangement could have been effected.

A detailed report containing the original data derived in response to questionnaires dealing with the intimacies of sexual behavior would, it is appreciated, be impossible at present. It is hoped, however, that investigators will arrange (as Hamilton seems to suggest that he has done) for their original material to be published or made available—anonymous still, but in complete detail—say, some fifty years hence. It is not impossible that some investigator might, even in those enlightened days, be sufficiently interested in sex behavior to perform a necropsy.

A further comment, which is pertinent to the publication of data, relates to the need for a complete, detailed statement of whatever data are given. It is impossible to pool data from different studies when, for example, one or another of the investigators classifies the ages at which certain forms of behavior first made themselves evident, as being above, at, and below a given age. The number, as well as the percentage, of cases should be stated, and for every age year-level, etc., separately.

In no case has the statistical reliability of a comprehensive sex-questionnaire been reported. To be able to do so probably would involve expensive repetition, and incur considerable resistance on the part of some subjects. It could be attempted. Indeed, for a group of items contained in his study, Chassell reports an average coefficient of reliability of seven-tenths. The more pertinent issue, however—namely, whether items of different context and implication (depending upon ability to recall, emotional conditioning, etc.) differ in their degree of reliability, and if so, in what way—has not been reported. Measures of reliability such as these, would, it is to be noted, still constitute a measure of reliability but only of conventional responses, and not necessarily of the actual springs of behavior (a problem discussed above).

Finally, if the observer can induce individuals other than those

in whom his questionnaire investigation is directly concerned to reply to the same questionnaire, or, if he can elicit information from all kinds of individuals, and then separate his cases into mutually exclusive categories, he will be able to check against the common logical fallacy of assuming that "antecedence-subsequence" relationships which obtain in the observed group necessarily are peculiar and significant to none other than that group. The fact that a questionnaire "seeks out" subjects and does not have to wait for them to select themselves, combined with the relative ease with which it may be answered, and the large number of cases which it may reach, if properly conducted, make this important control possible.

5. *The Laboratory Procedure*

Few cases; normal persons; situation strictly constant; observer constant; objective interpretation of results obtained by means of instruments; no control group; statistics incomplete.

Characteristic of this procedure are studies by Eagleson (9), and Hollingworth (21).

The laboratory procedure of studying sex behavior has not as yet been favorable to the investigation of large numbers of cases. It does, however, assure strictly constant observational conditions. The stimulus is provided by objective, constant instruments, such as mental test, or by physiological periodic changes in the body of the subject. The response is recorded in part equally objectively, in writing, or as smears on microscope slides. The introspection of subjects concerning, for example, fatigue, or the intensity of their sex desire, is, however, a factor of doubtful reliability. The interpretation of response is determined relative to established norms on a simple scale. So far no use of control groups has been reported. What statistics are available are satisfactorily treated, but, in view of the small number of cases, are of little significance. It is to be noted that this procedure differs but little from the laboratory procedure described in the preceding chapter.

* * * * *

How may we summarize this part? Apparently, so far as human sexual behavior is concerned, the method of personal report proffers at present the more productive method of investigation. Conceding its defects—most particularly that of unreliability due to human rationalization and forgetfulness—it still seems to yield data both more comprehensive and more self-revelatory than does the method

of direct observation. But the worth of the different procedures which subsume under this category appears to vary quite considerably from one to another depending upon the procedure and the investigation in question. Of the five procedures outlined, once more the "laboratory" constitutes the most precise, but again the least fecund. The "anecdotal," at the other extreme, though big with suggestions, in itself delivers little that may be termed shapely and conformable to scientific requirements. To determine the relative values of the remaining three procedures—apparently the most fruitful of all—is difficult. Questionnaire, case study, and psychoanalytic clinic—each seems to be little more than the complement of the other two. They rightly should constitute not three distinct procedures but three aspects of one. Seeking to evaluate each in turn, in terms of scientific method, it would appear that the psychoanalytic is the least satisfactory. It is as yet too unreliable, too personal, too interpretive, to warrant classification as a measuring and recording device. If its technique can be standardized it might turn out to be one of the most powerful instruments of scientific method. The questionnaire and case study remain. The strength of the former lies in its flexibility, its mobility, its extensiveness; but inherently weak is its validity. Case study, on the other hand, though relatively weak on the side of sampling, yields probably the most valid device for measuring inter-human relationships over a fairly extensive field. It is slower moving than is the questionnaire; but it is undoubtedly more effectual. The questionnaire, on the other hand, is probably the more reliable. Combining the questionnaire with the case study, as complementary functions of the same method, should yield the most productive as well as the most reliable results in the study of human sexual behavior.

REFERENCES

1. ACHILLES, P. S. The effectiveness of certain social hygiene literature. New York: Amer. Soc. Hygiene Asso., 1923. Pp. 116.
- 1a. ACKERSON, L. Children's behavior problems. Chicago: Univ. Chicago Press, 1931. Pp. xxi+268.
2. ADLER, A. The practice and theory of individual psychology. New York: Harcourt, Brace; London: Kegan Paul, 1924. Pp. 352.
3. BINGHAM, H. C. Sex development in apes. *Comp. Psychol. Monog.*, 1928, 5, No. 23. Pp. 166.
4. BRIFFAULT, R. The mothers. New York: Macmillan, 1927. (3 vols.) Pp. xix+781; xx+789; xv+841.
5. CHASSELL, J. O. The experience variables. Rochester, N. Y.: Univ. Rochester, 1928. Pp. 42.

6. DAVIS, K. B. Factors in the sex life of twenty-two hundred women. New York: Harpers, 1929. Pp. xx+430.
- 6a. DICKINSON, R. L., & BEAM, L. One thousand marriages. Baltimore, Md.: Williams & Wilkins. Pp. xxv+483.
7. DICKINSON, R. L., & PIERSON, H. H. The average sex life of American women. *J. Amer. Med. Asso.*, 1925, **85**, 1113-1117.
8. DUNLAP, K. Mysticism, Freudianism, and scientific psychology. St. Louis, Mo.: Mosby, 1920. Pp. 173.
9. EAGLESON, H. E. Periodic changes in blood pressure, muscular coordination, and mental efficiency in women. *Comp. Psychol. Monog.*, 1927, **4**, No. 20. Pp. 65.
10. ELLIS, H. Studies in the psychology of sex. Vol. 7. Philadelphia: Davis, 1930.
11. EXNER, M. J. Problems and principles of sex education. New York: Asso. Press, 1915. Pp. 39.
12. FERENCZI, S. The theory and technique of psychoanalysis. New York: Boni & Liveright, 1926. Pp. 473.
13. FREUD, S. Introduction to psychoanalysis. New York: Liveright, 1920. Pp. x+406.
14. GESELL, A. Infancy and human growth. New York: Macmillan, 1928. Pp. xvii+418.
15. HALL, G. S. Adolescence. New York: Appleton, 1904. (2 vols.) Pp. xx+591; 784.
16. HAMILTON, G. V. A study of sexual tendencies in monkeys and baboons. *J. Anim. Behav.*, 1914, **4**, 295-318.
17. ———. A research in marriage. New York: Boni, 1929. Pp. xiii+570.
18. HAMILTON, G. V., & MCGOWAN, G. What is wrong with marriage? New York: Boni, 1929. Pp. xxi+319.
- 18a. HARTWELL, S. W. Fifty-five bad boys. New York: Knopf, 1931. Pp. xvii+359.
19. HEALY, W. The individual delinquent. Boston: Little, Brown, 1915. Pp. 830.
20. HEALY, W., BRONNER, A. F., & BOWERS, A. M. The structure and meaning of psychoanalysis. New York: Knopf, 1930. Pp. xx+482+xxiv.
21. HOLLINGWORTH, L. S. Functional periodicity. New York: Teach. Coll., 1914. Pp. 101.
22. JACOBSON, E. Progressive relaxation. Chicago: Univ. Chicago Press, 1929. Pp. xiii+429.
23. JUNG, C. Analytical psychology. New York: Moffat, Yard, 1916.
24. KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insan.*, 1910, **67**, 37-96, 317-390.
25. KING, J. L. Menstrual records and vaginal smears in a selected group of normal women. *Carnegie Instit. Wash., Contrib. Embryol.*, 1926, **18**, No. 95.
26. KRAFT-EBING, R. V. Psychopathia sexualis. New York: Med. Art Agency, 1922. Stuttgart: Enke, 1924. Pp. vii+832.
27. LECKY, W. E. H. History of European morals. New York: Longmans, Green, 1869. (2 vols.)

28. LINDSEY, B. B. *The revolt of youth*. New York: Garden City Publ. Co., 1925. Pp. 364.
29. LIPSCHUTZ, A. *The internal secretions of the sex glands*. Cambridge: Heffer; Baltimore, Md.: Williams & Wilkins, 1924. Pp. xviii+513.
30. LUNDBERG, G. A. *Social research*. New York: Longmans, Green, 1929. Pp. 380.
31. MALINOWSKI, B. *Sex and repression in savage society*. New York: Harcourt, Brace, 1927. Pp. 299.
32. ———. *The sexual life of savages*. New York: Liveright, 1929. Pp. 500.
33. MARSHALL, F. H. A. *The physiology of reproduction*. New York: Longmans, Green, 1922. Pp. 770.
34. MEAD, M. *Coming of age in Samoa*. New York: Morrow, 1928. Pp. 297.
35. MILES, C. C., & Terman, L. Sex differences in the association of ideas. *Amer. J. Psychol.*, 1929, **41**, 165-206.
36. MOLL, A. *The sexual life of the child*. New York: Macmillan, 1913.
37. PEARL, R. *The biology of population growth*. New York: Knopf, 1925. Pp. 260.
38. PECK, M. W., & WELLS, F. L. On the psychosexuality of college graduate men. *Ment. Hygiene*, 1923, **7**, 697-714.
39. ———. Further studies in the psychosexuality of college graduate men. *Ment. Hygiene*, 1925, **9**, 502-520.
40. POPNOE, P. Eugenic sterilization in California. *J. Abn. & Soc. Psychol.*, 1929, **24**, 251-268.
41. PRESSEY, S. L. A group scale for investigating the emotions. *J. Abn. Psychol.*, 1921, **16**, 55-64.
- 41a. STEKEL, W. *Psychoanalysis and suggestion therapy*. New York: Moffat, Yard, 1923. Pp. xi+155.
42. THOM, D. A. *Everyday problems of the everyday child*. New York: Appleton, 1927. Pp. xiv+350.
- 42a. WECHSLER, D. The incidence and significance of fingernail-biting in children. *Psychoanal. Rev.*, 1931, **18**, 201-209.
43. WESTERMARCK, E. *History of human marriage*. (5th ed., 3 vols.) London: Macmillan, 1921.
44. WILLOUGHBY, R. R. The efficiency of short psychoanalysis. *J. Abn. & Soc. Psychol.*, 1931, **26**, 125-130.
45. [Various.] The conception of sexuality. (A symposium.) *Brit. J. Med. Psychol.*, 1925, **5**, 161-210.

55 Eustis Street

Cambridge, Massachusetts

L'ÉTUDE SCIENTIFIQUE DU COMPORTEMENT SEXUEL HUMAIN

(Résumé)

On peut considérer scientifique l'étude du comportement sexuel humain à mesure qu'elle se conforme aux exigences fondamentales de méthode employées ordinairement dans les procédés scientifiques. Dans cet article on essaie d'évaluer, en termes de ces exigences fondamentales ordinaires, la valeur des différents procédés employés actuellement comme instruments dans l'étude du comportement sexuel humain. On établit une différence entre l'observation directe et celle qui est possible seulement au moyen de la parole. Sauf dans les domaines de la recherche anthropologique et de l'étude des enfants, il paraît que l'observation directe n'ajoutera pas beaucoup à notre connaissance du comportement sexuel humain. On n'a pas encore développé une technique de l'observation directe laquelle est tout-à-fait satisfaisante au point de vue scientifique. Sous le titre Observation Indirecte, cependant, on fait attention surtout aux procédés étude des cas, psychoanalytique, et questionnaire. On les considère trois aspects de la même méthode plutôt que des procédés distincts et séparés. Le procédé psychoanalytique, une fois standardisé, pourrait se montrer un des plus puissants instruments de la méthode scientifique. Jusqu'au moment actuel, cependant, il n'est pas satisfaisant. Le procédé étude des cas est probablement le plus profond et possède le plus de valeur. Le questionnaire est probablement le plus compréhensif, le plus constant, et le plus flexible. On recommande une combinaison des trois procédés.

HARVEY

DAS WISSENSCHAFTLICHE STUDIUM DES MENSCHLICHEN
GESCHLECHTSVERHALTENS

(Referat)

Das Studium des menschlichen Geschlechtsverhaltens kann insofern als wissenschaftlich betrachtet werden, als es den grundsätzlichen methodischen Forderungen entspricht, die beim wissenschaftlichen Verfahren üblich sind. In dieser Arbeit wird der Versuch gemacht, auf Grund dieser herkömmlichen grundsätzlichen Forderungen, den Wert der verschiedenen, heute angewandten Verfahren, als Instrumente der Untersuchungen im Gebiete der Geschlechtsverhaltens zu bestimmen. Die direkte Beobachtung wird von derjenigen unterschieden, die nur durch das Mittel der Sprache möglich ist. Ausser im Gebiet der anthropologischen Untersuchungen und im Studium der Kinder scheint die direkte Beobachtung nicht viel beizutragen zu unserem Wissen über das Geschlechtsverhalten. Es ist bis jetzt noch keine wissenschaftlich hinreichende Technik der direkten Methode entwickelt worden. Unter dem Titel der indirekten Methode sind aber das Studium einzelner Fälle, das psychoanalytische Verfahren und die Methode des Fragebogens eingehend behandelt worden. Sie werden eher als drei Gesichtspunkte ein und derselben Methode denn als abgesonderte Verfahren gehalten. Das psychoanalytische Verfahren, wenn es einmal normiert sein wird, mag sich als eines der wirksamsten Instrumente der wissenschaftlichen Methode erweisen. Bis dahin aber ist sie ungenügend. Das Studium einzelner Fälle ist wahrscheinlich das gründlichste und zuverlässigste. Der Fragebogen ist wahrscheinlich das umfassendste und biegsamste Verfahren. Es wird eine Kombination der drei empfohlen.

HARVEY

A GROUP TEST FOR THE MEASUREMENT OF CRUELTY-COMPASSION: A PROPOSED MEANS OF RECOGNIZING POTENTIAL CRIMINALTY*¹

From the Department of Psychology, University of Chicago

JOSEPH WYMAN HAWTHORNE

PURPOSE OF THIS INVESTIGATION

It was the purpose of this study to devise a test that applied to a more restricted field than heretofore attempted in the study of criminal or delinquent tendencies and to lengthen the test in the hope of improving the reliability. It was hoped that a test could be devised which would have some predictive value so that early tendencies might be noted and treated. The field chosen was that which involved crimes against the person. Such a field seemed to be socially the most significant, and it was thought that if early tendencies toward this sort of criminal behavior could be detected in even a very small number of cases the test would be worth while.

NATURE OF THE TRAIT TESTED

Upon analysis it is readily seen that "crimes-against-the-person" is by no means a simple classification. The following is an attempt to make a classification of the mental tendencies which might predispose an individual to commit a violent crime against the person of another:

1. *Sadistic Tendencies.* Sadism is generally considered to be some form of perversion of the sexual impulse in which sexual gratification is obtained by inflicting pain. It exists in all stages from the

*Recommended by Harvey A. Carr, accepted for publication by L. L. Thurstone of the Editorial Board and received in the Editorial Office, April 27, 1931.

¹Whatever of value may exist in this thesis is due to the patient, kindly, and keen criticisms of Professor Harvey A. Carr, whose interest and encouragement were responsible for the completion of the work. Thanks are also gratefully extended to Mr. Willis Clark of the Psychology Department of the Los Angeles city schools for ideas drawn from his "Delinquency Index Scale," to Mr. and Mrs. Z. R. DeArmond and Miss Helen Baum for the administration of a large number of the tests, and to Mr. Paul Kemske for valuable aid in clerical and statistical work.

type of individual who enjoys teasing another to the type who will not hesitate to commit murder at the slightest opportunity.

2. *Pathological Irritability.* In this type of individual an intense anger is aroused by a relatively insignificant and often unrecognizable stimulus. It is during such a state that many of the so-called justifiable homicides occur, as well as many other types of crime with vengeance as a motive.

3. *Pathological Unemotionality.* A person with this trait is said to be unable to experience such feelings as grief, pity, compassion, and the like, which might tend to act as deterrents to crime. Some of the famous "thrill" murders are supposed to have been committed by persons of this type.

4. *Delusional Trends.* The individual in this classification who is apt to commit a violent crime is one who is suffering from delusions of persecution, however aroused. His crimes are committed in self-defense against some person whom he imagines has the intention to harm him. Most of the murders of public officials have been committed by people of this type.

5. *Hallucinations.* Many murders have been committed by persons whose defense was that God or an angel told them to do it. These people are undoubtedly telling the truth, and a number of murders can be accounted for on the basis of the hallucinations of insanity, drug addiction, alcohol addiction, and the like.

6. *Criminalism.* Some psychiatrists insist on a special classification of a mental disorder variously called criminalism, moral insanity, moral imbecility, or the like, which is usually listed under the constitutional psychopathic states. This type of individual seemingly is unable to acquire any ethical discriminatory ability although other functions seem normal. Whether or not such a constitutional disorder actually exists, it is at least worth listing as a possible cause of violent crimes.

7. *Mental Deficiency.* With the introduction of mental testing it was thought that mental deficiency was one of the major causes of crime. With further study this seems not to be the case. Nevertheless, there is a much greater chance of a crime being committed by a person unable to foresee the consequences of the act than by a person of greater intelligence, other things being equal. In spite of the fact that perhaps the intelligent criminal is the one who is not caught, there seems to be a much higher percentage of feeble-minded among the criminal population than among normal groups. (Vari-

ously estimated as from 8-50% among criminal population and 2-3% among normal population.)

8. *Hypersuggestibility*. Whether or not there is a definite trait of suggestibility it is highly probable that Macbeth was not the only person who had been goaded into a crime by the insistence of another. Whether with or without cause, the suggestive influence of the vivid portrayal of crimes on the screen has often been commented upon.

Whatever may be the mental twist which causes an individual to attack another, it does seem that there is a general trait, usually called cruelty, which predisposes the person possessing it toward overt acts of cruelty. It would also be expected that this trait exists in varying degrees in different individuals ranging from cruelty at one extreme to compassion on the other. It was in an attempt to rate individuals on this scale of cruelty-compassion that this test was devised.

As to the question as to whether the tendencies which make for crimes of violence are native or acquired, this study is not concerned. What is insisted upon is that before the adult indulges in overt criminal behavior he must first have certain interests and attitudes which will predispose him to that behavior. The criminal sadist, for example, who as an adult commits murder, as a child in all probability enjoyed torturing animals or other children. The gangster killer is assumed to have had a different way of evaluating things as a youth than the person who studiously avoids any sort of cruelty.

CONSTRUCTION OF THE TEST

In an attempt to get at these interests and attitudes the following test² was constructed, which is largely self-explanatory. The subject is asked to rate his preference for each of the items in each section of the test in order from one to five. There are 31 sections, each of which contains one item which *a priori* was considered to be of sadistic (in its broadest sense) import. It is assumed that the attitude of the subject toward this one significant item in relation to the others will give some indication of his reaction tendencies.

ADMINISTRATION OF THE TEST

The test may be passed out to as large groups as can be conveniently handled. The subjects, who are elementary or high-school

²Appended to end of this paper.

students, merely read the directions and act accordingly. There is no time limit on the test, and the subjects are invited to ask about any of the items that they do not understand. It is announced as a test of student preferences, or as a game, or what-not, as long as the real purpose of the test is not revealed to the subjects. The administration is thus a simple matter as it does not require any special supervision and can be given out at odd times to school pupils who can finish part or all of the test at a sitting.

SCORING OF THE TEST

The tests are scored by drawing a circle around the significant or key item in each of the sections, and by simply adding up the numbers which appear in the circles, the other items being disregarded. Thus, if the significant item is rated a "five" in each case the score will be 155 which, if the test means anything, means that the subject shows tendencies which are the opposite of cruel. On the other hand, if the significant items are all rated a "one" it would tend to show that the subject had a condoning attitude toward cruelty. Thus the lower the score the higher is the degree of tolerance toward acts of cruelty and supposedly the more the tendency to overt acts of cruelty. The higher the score the greater the degree of compassion or the less the tendency toward overt acts of cruelty.

STANDARDIZATION OF THE TEST

The tests were given to a group of 126 students of regular junior and senior high schools who ranged in age from 10 to 20 years. This was used as a normal or standard group. With the results obtained from the normal group were compared the results from slightly over 300 students in special schools for juvenile delinquents and from 29 patients of a state insane hospital.

Special No. 1 was a high school of a large California system whose enrollment was largely, though not entirely, made up of discipline cases from other high schools of the system.

Special No. 2 was a state school for juvenile delinquents whose enrollment was made up of cases committed by the courts.

Special No. 3 was the Chicago Cook County School for Boys where the enrollment is also made up of court cases.

Special No. 4 was the Kankakee State Hospital for the Insane.

Table 1 shows the results obtained. Column 1 shows the school, Column 2, the number of cases, Column 3 the means, Column 4

TABLE 1

	Number of cases	Mean	<i>P.E.</i> (mean)	Standard deviation	<i>P.E.</i> (sigma)
Normal	126	119.10	0.88	14.68	.62
Special No. 1	178	108.17	0.81	16.16	.58
Special No. 2	71	105.30	1.32	16.50	.93
Special No. 3	58	102.17	1.56	17.67	1.113
Special No. 4	29	119.00	2.19	17.52	1.56

the *P.E.* of the means, Column 5 the standard deviation from the mean, and Column 6 the *P.E.* of the *S.D.*

These differences were then compared to see whether or not they were statistically reliable with the results as shown in Table 2. Column 1 shows the special schools each grouped with the normal high school. Column 2 shows the means, Column 3 the *P.E.* of the means, Column 4 the differences between the means of each special school and the normal high school, and Column 5 shows the difference divided by the *P.E.* of the difference.

TABLE 2

	Mean	<i>P.E.</i> (mean)	Diff.	<i>P.E.</i> (diff.)	Diff. <i>P.E.</i> (diff.)
Normal	119.10	.88			
Special No. 1	108.17	.81	10.93	1.19	9.18
Normal	119.10	.88			
Special No. 2	105.30	1.32	13.80	1.59	8.67
Normal	119.10	.88			
Special No. 3	102.17	1.56	16.93	1.79	9.46
Normal	119.10				
Special No. 4	119.00				

RELIABILITY OF THE TEST

The reliability of the test was determined, first, by correlating the first half against the second half and correcting by means of the Spearman-Brown formula, and, secondly, by correlating the odd against the even items and correcting as before. In addition to this, the tests were given twice to 21 college freshmen and the correlation between the two administrations of the test was obtained. This result is appended to the following table for what it is worth.

	<i>N</i>	<i>R</i>	<i>P.E. (R)</i>
First half <i>vs.</i> second half	261	.71	.02
Odd <i>vs.</i> even	234	.66	.02
Two administrations	21	.93	.02

INDEX OF RELIABILITY

The correlation between the obtained scores and their corresponding "true" scores was obtained by the formula:

$$R \text{ (obtained score)} = \sqrt{R12}$$

which in this case becomes:

$$R \text{ (obtained score)} = \sqrt{.71} = .84$$

PROBABLE ERROR OF MEASUREMENT

In order to obtain the probable error of any individual score the following formula was used:

$$P.E. \text{ (meas.)} = .6745 \sigma \sqrt{1-R(12)}$$

which in this case becomes:

$$P.E. \text{ (meas.)} = .6745 \times 17.67 \times \sqrt{1-.71} = 6.42$$

Thus the chances are even that the score of any individual lies between 6.41 points above or below his obtained score.

VALIDITY OF THE TEST

A question which at once presents itself in a test of this kind is whether the subjects are reacting on the basis of some real attitude, the measure of which the test results show, or whether they are able to appreciate the object of the test and react on the basis of intelligence or perhaps mere chronological age. If this were the case it would seem (*a*) that the subjects would be able to tell what the test was designed to measure as soon as they had taken it, (*b*) that the scores would be uniformly high, and (*c*) that there would be at least a positive if not a high positive correlation between test results and intelligence and between test results and chronological age.

As a check on the first possibility practically all subjects taking the test were asked informally concerning the nature of the test. In no case was the correct diagnosis made.

As a check on the second possibility both IQ's and chronological

ages were correlated with test results as shown in the following table:

	Number of cases	<i>R</i>	<i>P.E. (R)</i>
IQ—test score	101	— .02	.07
Age—test score	325	— .03	.04

The really adequate method of getting at the validity of this kind of test would be to study a group of individuals taking it, and to carry the study through a period of ten or twenty years, correlating subsequent behavior with the original test results. At the present time such a procedure is, of course, impossible. In lieu of such a program, information was obtained on a group of individuals who had been admitted to special institutions to see if there were any relation between test score and such behavior as had already been observed. This information was obtained from case histories, commitment records, interviews with parents and teachers. First a group, all of whom obtained scores below the mean of the special schools, was studied in this manner. Here follows a summary of the evidence obtained.

Ev. Tu. Age 49. Test score 99. Committed to Kankakee State Hospital for wanting to kill his brother. Released and worked a year as butcher. At time of his second commitment he had a seizure in which he jumped into a pen and tried to kill a steer. While in hospital he picked up a pair of scissors and tried to attack an attendant.

Ew. Je. Age 30. Test score 105. Committed to Kankakee State Hospital after getting into a fight at the home of his sister and becoming violent with the police. Patient had persecutory delusions and believed people were going to kill him.

Ot. Ha. Age 20. Score 102. Committed to Kankakee State Hospital after becoming violent and attacking his father again and again in a fit of rage. While in hospital took picture from the wall and attacked another patient.

Pa. W. Age 23. Score 104. Started a fight in a cafe. Was arrested and sent to psychopathic and finally committed to the State Hospital.

Jo. Kl. Age 34. Score 100. Committed to Kankakee with a psychosis, but with nothing in his case history to indicate troublesome nature. Hospital records, however, shows record of treatment for slight wounds received while in fights with another patient. Was in two fights in the month previous to examination of his record.

Ha. Br. Age 18. Score 98. Committed to special school for theft. Reported as a very trying pupil in the classroom and as a bully on the playground.

Al. Da. Age 15. Score 91. Committed to special school for truancy. Described as irritable, quick-tempered, often in fights with other boys.

Be. Ep. Age 15. Score 54. Abusive, uncooperative, with spells of violence. Committed to special school for being impossible to handle in the classroom. Refuses to do ordinary school work and is interested only in cheap detective thrillers.

Ca. Hu. Age 15. Score 90. Committed to special school for continual truancy. No record of cruel behavior in school. Interview with the parents, however, brought out the fact that the subject was continually torturing the other three children all of whom were very bitter against him.

Jo. Mi. Age 16. Score 82. Committed to special school because of fighting, bullying, and constantly provoking quarrels.

Fe. Be. Age 17. Score 93. Committed for truancy. While at special was noted as a bully with the younger boys whom he tormented on the playground.

Be. Iv. Age 17. Score 103. Committed to special for being incorrigible in the classroom, tormenting all of his teachers for no known reason.

Hu. Ca. Age 16. Score 93. Committed to special because he would get into a fight on the slightest provocation.

R. Gr. Age 13. Score 95. Described by his special school teachers as being "just naturally ornery." Continually in trouble with other boys.

Lo. Ke. Age 17. Score 96. Described as mean, troublesome, and fond of tormenting other boys.

Jo. Mc. Age 15. Score 95. Troublesome in classroom. Once stabbed another boy with a knife.

Wa. Vi. Age 16. Score 74. Bully on the school ground. Continually in fights. Other boys are afraid of him.

Li. Tr. Age 17. Score 78. Committed for general incorrigibility, fighting, meanness, and for tormenting both teachers and pupils.

Ma. Fa. Age 14. Score 105. Will not mingle with boys of his own age. Likes to bully younger boys.

Ja. Dr. Age 16. Score 86. Described by special school teachers as mean, irritable, and a bully.

Ch. Au. Age 16. Score 98. Mean, sneaky, and stubborn. Has given trouble to all of his teachers.

Em. El. Age 16. Score 99. Committed to special school for having in his possession a quantity of dynamite with which he intended to blow up his enemies.

To. Ro. Age 16. Score 86. Considered "psychopathic" by his special school teachers because of his meanness and cruelty.

Pa. Me. Age 15. Score 103. Described by special school teachers as very mean.

El. Ba. Age 15. Score 102. Bully. Attacks other boys without any provocation.

Ep. Ca. Age 15. Score 83. Full-blooded Indian which is reason given by his special school teachers for his meanness and cruelty.

Al. Ga. Age 17. Score 89. Described as "hard-boiled" and a bully by his teachers.

Fr. Oa. Age 15. Score 89. Already considered "dangerous" by his teachers who freely predict that he will end up in serious trouble.

La. Sh. Age 16. Score 100. Intensely hates most of the other boys in the school who call him "Gopher" because of his protruding front teeth. Becomes violently angry and attacks other when teased.

Ha. Ph. Age 15. Score 90. Described as mean, irritable, and stubborn.

Ru. Tu. Age 15. Score 73. Described as mean, ornery, and continually picking fights with other boys who are smaller than he.

Al. De. Age 15. Score 103. Has periodic spells of meanness and irritability.

La. O'T. Age 15. Score 73. Known at the special school as the "Tough Irishman." Always looking for a fight and often finding it.

Ke. Ev. Age 15. Score 95. Has spells of meanness and irritability.

A group of those special-school students who made high scores was studied in the same manner. (Those normal high-school subjects making high scores of course showed no evidence of cruel traits as would be expected.) The question arises, however, as to whether the low scores might not be indicative of general criminal tendencies. Accordingly, those students from special schools who made scores above the mean of the special students were studied. Although this group has already committed such crimes as stealing, forgery, counterfeiting, and the like, there was no evidence of such tendencies toward cruelty as were exhibited in the lower group previously surveyed.

It is considered significant that the cases of cruelty were found to be in the lower group of scores. While it is true that within the scope of this study some of the lower scores failed to show any record of cruel behavior, it must be remembered that, in the first place, complete information was very often unobtainable and, in the second place, many of the subjects were too young to have had much chance to exhibit the trait.

RELATIVE VALUE OF THE INDIVIDUAL TESTS IN THE BATTERY

In order to determine the relative value of the individual tests in the battery a group of scores were selected at random from the upper quartile of the entire group and compared with a similar group in the lower quartile. There were 56 cases in each group. All of the upper group were from normal schools and all but six of the lower group were from special schools. The method of comparison was as follows:

1. The average performance of the 56 individuals in the low group in each of the 31 tests in the battery was determined.
2. The average performance of the 56 individuals in the high group in each of the 31 tests was determined.
3. The difference between each of the 31 means was determined.
4. For each of the 31 tests the percentage of the upper group making scores equal to or exceeding the mean of the lower group was determined.

The higher these percentages, the greater is considered the value of the individual tests.

These results are summarized in Table 3. It will be seen that in

TABLE 3
PROGNOSTIC VALUE OF SEPARATE TESTS IN THE BATTERY

Test No.	Mean score low group	Mean score high group	Difference in means	Percentage of high group above mean of low group
1	2.473	4.182	1.709	92.7
2	3.875	4.625	.750	94.6
3	3.429	4.911	1.482	98.2
4	3.161	4.446	1.285	89.3
5	3.536	4.518	.982	85.7
6	2.945	3.500	.555	85.7
7	1.673	4.268	2.595	98.2
8	3.182	4.500	1.318	85.7
9	1.945	3.964	2.019	98.2
10	3.782	4.786	1.004	94.6
11	2.509	5.000	2.491	100.0
12	2.255	3.822	1.567	83.9
13	3.125	4.857	1.732	98.2
14	3.164	4.518	1.354	85.7
15	2.232	4.357	2.125	85.7
16	1.964	3.946	1.982	91.1
17	3.482	4.709	1.227	92.7
18	2.109	4.786	2.677	98.2
19	2.818	4.473	1.655	94.5
20	1.704	4.145	2.441	90.9
21	1.685	3.643	1.958	91.1
22	1.782	3.576	1.794	92.9
23	3.236	4.643	1.407	87.5
24	2.000	3.411	1.411	98.2
25	2.000	4.393	2.393	89.3
26	2.446	3.857	1.411	78.6
27	3.679	4.911	1.232	98.2
28	3.400	4.491	1.091	89.1
29	3.564	4.566	1.002	86.8
30	3.691	4.528	.837	88.7
31	3.545	3.852	.307	74.1

TABLE 4
MEANS AND *S.D.*'s, USING ALL 31 TESTS

	No. of cases	Mean	<i>P.E.</i> (<i>Av.</i>)	Standard Deviation	<i>P.E.</i> of (sigma)
Normal	136	118.24	1.02	17.65	.721
Special No. 1	112	111.35	1.06	16.70	.75
Special No. 2	69	107.75	1.45	17.80	1.02

17 out of the 31 tests the percentages are above 90, 11 are between 85 and 90; one between 80 and 85; one between 75 and 80; and one between 70 and 75.

In order to determine whether the efficiency of the test could be improved by leaving out those tests having a low prognostic value, the data in Tables 1 and 2 were recalculated after excluding certain tests from the total. The following tables show the results obtained.

Table 4 shows the means and the standard deviations for the normal and two special schools with their probable errors and the difference between the means. In this table the total score of all 31 tests was used. (The fact that these results are slightly different from Table 2 is due to the fact that a somewhat different but nevertheless unselected group was used for this part of the study. This is due to the fact that this part of the study was undertaken later and not all of the original tests were available.)

TABLE 5
MEAN AND DIFFERENCE IN MEANS, USING ALL 31 TESTS

	Mean	<i>P.E.</i> of Mean	Diff. (<i>Av.</i>)	<i>P.E.</i> of (Diff.)	Diff. (<i>Av.</i>) <i>P.E.</i> (Diff.)
Normal	118.24	1.02	6.89	1.47	4.68
Special No. 1	111.35	1.06			
Normal	118.24	1.02	10.49	1.77	5.91
Special No. 2	107.75	1.45			

TABLE 6
MEANS AND *S.D.*'s, LEAVING OUT TEST NO. 31

	No. of cases	Mean	<i>P.E.</i> (<i>Av.</i>)	Standard Deviation	<i>P.E.</i> of (sigma)
Normal	136	114.50	.989	17.10	.699
Special No. 1	112	107.70	1.05	16.65	.748
Special No. 2	69	104.01	1.37	16.90	.99

TABLE 7
MEANS AND DIFFERENCE IN MEANS, LEAVING OUT TEST NO. 31

	Mean	P.E. of (mean)	Diff. (Av.)	P.E. of (Diff.)	Diff. (Av.) P.E. (Diff.)
Normal	114.5	.99	6.8	1.44	4.72
Special No. 1	107.7	1.05			
Normal	114.5	.99	10.49	1.70	6.17
Special No. 2	104.01	1.37			

TABLE 8
MEANS AND S.D.'s, LEAVING OUT TESTS NOS. 31 AND 26

	No. of cases	Mean	P.E. (Av.)	Standard deviation	P.E. of (sigma)
Normal	136	111.54	.96	16.75	.685
Special No. 1	112	104.50	1.03	16.20	.728
Special No. 2	69	100.68	1.335	16.45	.96

TABLE 9
MEANS AND DIFFERENCE IN MEANS, LEAVING OUT TESTS NOS. 31 AND 26

	Mean	P.E. of Mean	Diff. (Av.)	P.E. of (Diff.)	Diff. (Av.) P.E. (Diff.)
Normal	111.54	.96	7.04	1.40	5.02
Special No. 1	104.50	1.03			
Normal	111.54	.96	10.86	1.64	6.62
Special No. 2	100.68	1.34			

TABLE 10
MEANS AND S.D.'s, LEAVING OUT TESTS NOS. 31, 26, AND 12

	No. of cases	Mean	P.E. (Av.)	Standard deviation	P.E. of (sigma)
Normal	136	108.15	.93	16.18	.66
Special No. 1	112	101.74	1.02	16.10	.72
Special No. 2	69	97.58	1.28	15.85	.91

TABLE 11
MEANS AND DIFFERENCES IN MEANS, LEAVING OUT TESTS NOS. 31, 26, AND 12

	Mean	P.E. of Mean	Diff. (Av.)	P.E. of (Diff.)	Diff. (Av.) P.E. (Diff.)
Normal	108.15	.93	6.41	1.38	4.64
Special No. 1	101.74	1.02			
Normal	108.15	.93	10.57	1.58	6.69
Special No. 2	97.58	1.28			

TABLE 12

MEANS AND *S.D.*'s, LEAVING OUT TESTS Nos. 31, 26, 12, 5, 6, 8, 14, AND 15

	No. of cases	Mean	<i>P.E.</i> (Av.)	Standard deviation	<i>P.E.</i> of (sigma)
Normal	136	89.01	.81	14.10	.576
Special No. 1	112	83.09	.833	13.085	.588
Special No. 2	69	79.83	1.09	13.50	.779

TABLE 13

MEANS AND DIFFERENCE IN MEANS, LEAVING OUT TESTS
Nos. 31, 26, 12, 5, 6, 8, 14, AND 15

	Mean	<i>P.E.</i> of Mean	Diff. (Av.)	<i>P.E.</i> of (Diff.)	Diff. (Av.) <i>P.E.</i> (Diff.)
Normal	89.01	.81	5.92	1.16	5.10
Special No. 1	83.09	.83			
Normal	89.01	.81	9.18	1.36	6.75
Special No. 2	79.83	1.09			

In order to determine what effect the cutting-out of certain items of the test had on the reliability, the reliability of the test was re-computed after removal of certain tests. For this purpose one hundred unselected cases were taken from the normal high school and the first half correlated against the second half and the result corrected by the Spearman-Brown Formula. The results are shown on Table 14. Column 1 shows the numbers of the tests used in the correlation, Column 2 the correlation obtained, Column 3 the *P.E.* of *R*, and Column 4 the numbers of the tests excluded.

As a means of further checking up on the relative value of the individual tests, those tests of a similar nature were grouped together and studied from this angle. Twelve of the individual tests involve

TABLE 14

N = 100

Number of tests used	<i>R</i>	<i>P.E.</i> (<i>R</i>)	Tests excluded
30	.80	.02	Nos. 31
28	.86	.02	Nos. 31, 26, 12
24	.83	.02	Nos. 31, 26, 12, 5, 6, 8, 14
22	.77	.03	Nos. 31, 26, 12, 5, 6, 8, 14, 29

TABLE 15

Percentage of upper group exceeding mean of lower group	"Watching" tests	"Owning" tests	"Judgment" tests	"Acting" tests
95-100	2	5	0	1
90-95	7	1	1	0
85-90	2	5	3	1
80-85	0	1	0	0
75-80	1	0	0	0
70-75	0	0	1	0
	—	—	—	—
	12	12	5	2

the watching of some act of cruelty, 12 involve the ownership of some means of inflicting cruelty, 5 involve the judgment of acts of cruelty in others, and 2 involve the commission of overt activity.

These results are shown in Table 15. The figures in the left-hand column refer to the percentages by which the upper group, studied in Table 3, exceeds the mean of the lower group. The other figures refer to the number of the tests of the various types which fall within these percentages.

USE OF TEST

It is proposed that this test shall be used on children, more especially boys, of junior- and senior-high-school age. It can be given in large groups without special instruction or timing. Whether the test has any predictive value or not, it should be of value for some of the modern attempts at character education, since a subject making a score below about 110 seems from the very nature of the test to be in need of some mental readjustment. If the test does have some predictive value, and the results seem to show that it does, it is offered as a mean of recognizing the incipient gang killers and murderers of various kinds before they commit their acts.

SUMMARY AND CONCLUSIONS

1. The results seem to show that the test furnishes the possibility of rating individuals along a cruelty-compassion continuum.
2. The results seem to show that those individuals making low scores in the test are more predisposed to crimes-against-the-person than are those individuals making higher scores.
3. The test results have zero correlation with both age and intelligence.

4. Certain of the test items have greater differentiating value than others.

5. The greatest differentiating value seems to result from the exclusion of tests numbers 31, 26, 12, 5, 6, 8, 14, and 15.

6. The greatest reliability is obtained by the exclusion of tests numbers 31, 26, and 12 ($R=.86$; $P.E. (r)=.02$).

7. No valid conclusions could be drawn as to the comparative efficacy of test categories: "watching," "owning," "judgment," or "acting."

8. The probable error of an individual score is 6.4 (When optimum combination of tests used, i.e., by excluding tests numbers 31, 36, and 12, the $P.E. (meas.)$ is 3.4.)

HISTORICAL SURVEY OF PREVIOUS WORK

One of the earliest tests in this field was devised by Dr. G. C. Fernald (5) which required the subjects to rate a series of crimes in order from the least to the most serious. His norm was determined from the ratings of a group of legal and scientific men. When the test was given to delinquent and criminal groups, he found wider variations from the norm than he found in the law-abiding groups.

W. W. Clark (4) devised a battery of seven tests, each of which consisted of rating seven crimes in the order of their seriousness. He also standardized his test by obtaining ratings from a group of professional people. When it was given to delinquent subjects, he found marked deviations from the ratings of the professional group. This was known as the Whittier Scale for Grading Juvenile Offenses.

Lask (8) in Germany, using one list of crimes in much the same manner as Fernald, had individuals rank the crimes in the order of their seriousness and obtained results similar to those of Fernald.

Raubenheimer (9), as a part of a larger test, used two forms taken directly from the Whittier Rating Scale in each of which the ten crimes were rated in the order of their seriousness.

Brogan (2) used the method of rating offenses in order to study the attitude of university students toward the 16 worst practices of university people.

Thurstone (10) used the method of paired comparisons in order to study the attitude of the group toward 19 major offenses and was able to rate the seriousness of the crimes on an absolute scale. In this study, as in the study of Brogan, the objective was a measure of group

attitude toward crimes rather than a study of individual relationships.

The foregoing studies are those which use rating methods as a study of either crimes or criminality. Cady (3), Raubenheimer (9), Allport (1), Hartshorne and May (6) and Hartshorne, May, and Welty (7) have made excellent surveys on the field of personality traits as a whole.

APPENDIX—THE TEST

NAME _____ AGE _____ GRADE _____
 SCHOOL _____ FATHER'S OCCUPATION _____
 DATE _____ WHAT DO YOU WANT TO BE WHEN
 YOU ARE OUT OF SCHOOL _____

TEST I

Suppose you had a chance to witness one of the sporting events listed below. Put a "1" after the one that you would like best to see and a "5" after the one that you would care least about seeing and rate the rest accordingly.

- a. A game of chess between two experts for the championship of the state. ☐
- b. A game of football between two of the largest colleges of the country. ☐
- c. A prize-fight between two contenders for the world's championship. ☐
- d. A track meet at the Olympic Games. ☐
- e. A grudge duel between two of Europe's most expert swordsmen. ☐

TEST II

Following is a list of the offenses committed by several boys of a certain high school. Put a "1" after the offense that you would condemn the least and a "5" after the one that you would condemn the most and rate the rest accordingly.

- a. Took a battery out of his father's car and sold it, using the money for shows and cigarettes. ☐
- b. Stole an English composition that another boy had written and handed it in for his own. ☐
- c. Jabbed a five-year-old girl with his knife until she was bleeding in several places. ☐
- d. Forged his mother's name to a check and cashed it. ☐
- e. Persistently played hookey from school and got a job in a factory. ☐

Suppose that you were to be given your choice of one article out of each of the following groups. The value of each article in a group is the same. Assuming that you were not already supplied with any of these things,

mark with a "1" the thing in each group that you would most rather have and with a "5" the thing in each group that you care least about and rate the rest accordingly.

TEST III

- a. sled
- b. dagger
- c. watch
- d. camera
- e. skates

TEST IV

- a. trapeze
- b. fencing set
- c. ice skates
- d. gold ring
- e. chemistry set

TEST V

- a. set of garden tools
- b. stiletto
- c. box of candy
- d. baseball glove
- e. microscope

TEST VI

- a. chess set
- b. sling shot
- c. carton of cigarettes
- d. deck of cards
- e. stamp album

TEST VII

- a. telescope
- b. golf clubs
- c. rifle
- d. pen and pencil set
- e. watch chain

TEST VIII

- a. hatchet
- b. pup tent
- c. Erecto set
- d. adventure books
- e. electric motor

TEST IX

- ☐ a. flashlight
- ☐ b. drawing set
- ☐ c. loaded dice
- ☐ d. hunting knife
- ☐ e. flute

TEST X

- ☐ a. rubber boots
- ☐ b. sweater
- ☐ c. black jack
- ☐ d. stamp album
- ☐ e. Eversharp pencil

TEST XI

- ☐ a. radio
- ☐ b. camping outfit
- ☐ c. set of story books
- ☐ d. revolver
- ☐ e. tennis racquet

TEST XII

- ☐ a. new hat
- ☐ b. cuff links
- ☐ c. dog
- ☐ d. boxing gloves
- ☐ e. new shoes

TEST XIII

- ☐ a. brass knuckles
- ☐ b. baseball
- ☐ c. whistle
- ☐ d. fountain pen
- ☐ e. brief case

TEST XIV

- ☐ a. pair of rabbits
- ☐ b. tobogganing sled
- ☐ c. dissecting set
- ☐ d. study lamp
- ☐ e. camping outfit

TEST XV

Suppose that through illness or for some such causes you are confined in such a manner that there is nothing whatever that you can amuse yourself with but to read and that the books listed below are the only ones available. Put a "1" after the one that you would like most to read and a "5" after the one that you would care least about, and rate the rest accordingly.

- a. "Confessions of a Russian Bluebeard." A Russian criminal while awaiting execution writes in detail of how he tortured and murdered fifty-three women and children and of how he escaped capture for so many years. ☐

b. "From Forest to School-room." The story of paper, and the many and interesting processes that it goes through in coming from the forest, through the various stages until it ends in the schoolroom in the form of a book. ☐

c. "The Voyage of the Ajax." The adventures encountered by a ship on a voyage to strange parts of the world in the quest for new specimens of plant life for a large museum. ☐

d. "Pioneer Days." The story of a covered wagon caravan making the trip to California in the days of the gold rush and of a boy who afterwards became the governor of California. ☐

e. "The Life and Work of Eugene Thomas." The story of a great scientist who gave his life to a study of human disease and by his careful and painstaking work saved the lives of millions although he lost his own. ☐

In each of the following groups of sports, put a "1" after the one that you like best to watch and a "5" after the one that you care least about watching and rate the rest accordingly.

TEST XVI

- a.* football
- b.* prize-fighting
- c.* baseball
- d.* basketball
- e.* track

TEST XXI

- ☐ *a.* pool ☐
- ☐ *b.* rifle-practice ☐
- ☐ *c.* dancing ☐
- ☐ *d.* ice skating ☐
- ☐ *e.* bowling ☐

TEST XVII

- a.* parachute-jumping
- b.* horse-racing
- c.* ski-jumping
- d.* dog-fighting
- e.* yachting

TEST XXII

- ☐ *a.* swimming ☐
- ☐ *b.* hunting ☐
- ☐ *c.* hiking ☐
- ☐ *d.* canoeing ☐
- ☐ *e.* bicycling ☐

TEST XVIII

- a.* sword fighting
- b.* surf-riding
- c.* gym work
- d.* archery
- e.* motor-boating

TEST XXIII

- ☐ *a.* cock-fighting ☐
- ☐ *b.* dog racing ☐
- ☐ *c.* polo ☐
- ☐ *d.* tobogganing ☐
- ☐ *e.* roller skating ☐

TEST XIX

- a.* wrestling
- b.* tennis
- c.* fencing
- d.* soccer
- e.* golf

TEST XXIV

- ☐ *a.* horseback-riding ☐
- ☐ *b.* chess ☐
- ☐ *c.* dice throwing ☐
- ☐ *d.* billiards ☐
- ☐ *e.* revolver practice ☐

TEST XX

- a.* rugby ☐
- b.* cricket ☐
- c.* handball ☐
- d.* ice hockey ☐
- e.* bull-fighting ☐

TEST XXV

Suppose that you were to be given an outing at the state capitol with special permission to attend anything that you liked, and you found all the events listed below occurring at the same hour. Mark with a "1" the

thing that you would like most to see and with a "5" the thing that you would care least about and rate the rest accordingly.

a. The legislature is to vote on whether or not that state shall refuse to abide by a recently adopted amendment to the federal constitution. ☐

b. The governor is to give out an interview to newspaper men which is of unusually great interest and which all the newspapers of the country will feature. ☐

c. Two famous criminals are to be executed by placing them in a glass chamber and injecting poisonous gases, a new method which was violently opposed because it was said to be unnecessarily cruel. ☐

d. The secretary of state is to meet with several of the biggest business men of the country to discuss the incorporation of a new business concern that is to be the biggest firm in the country. ☐

e. The highway commissioner is to dedicate formally a link of pavement that connects both coasts in the first continuous concrete auto highway in the country. ☐

TEST XXVI

Suppose that one of the following motion picture films is to be shown at your school. Put a "1" after the one that you would like best, and a "5" after the one that you think you would least enjoy and rate the rest accordingly.

a. "The High C's." A picture showing the entire process of how violins are made from the picking and testing of the wood to the final testing for tonal qualities. ☐

b. "From Pigs to Pork." A picture showing the handling and production of meat from the farm to the table. ☐

c. "Higher Learning." A picture showing the things that a student pilot has to know in handling an airplane above the clouds. ☐

d. "The Town Cut-up." A surgeon performs several major operations before the camera showing the details of surgical technique on actual patients. ☐

e. "Hot Stuff." Modern methods of fighting fires of various kinds and an actual picture of several large fires in which the various kinds of apparatus are used. ☐

TEST XXVII

Suppose that for some reason or other you are forced to work for a packing house and that you may have the choice of the jobs listed below, all of which carry the same pay and are of equal difficulty and offer the same opportunities for advancement. Put a "1" after the one that you would like best and a "5" after the one that you would care least about having and rate the rest accordingly.

a. Feeding and caring for the live stock as it is received from the farm. ☐

b. Working a lever that plunges a knife into the heart of each animal as it comes up a runway. ☐

c. Keeping records of the loaded freight cars, keeping track of when they are shipped, where they go, and the like. ☐

d. Weighting in the stock as it is received and keeping records of weights. ☐

e. Running a machine that pastes labels on tin cans. ☐

TEST XXVIII

Below are listed some of the charges that came before a certain court one morning. Disregarding any legal consequences of these acts, mark with a "1" the one that you would the least condemn and with a "5" the one that you would the most condemn and rate the rest accordingly.

- a. Held up a man at the point of a gun and forced him to give up his watch and purse containing \$350. ☐
- b. Keeper of an elaborate and richly furnished gambling house whose profits averaged about a thousand dollars a day. ☐
- c. Attempted to organize a strike among the workers of a large factory. ☐
- d. With some accomplices had assaulted and beaten into insensibility an enemy and left him bleeding on the sidewalk. ☐
- e. Had been boisterously shouting and singing in a quiet neighborhood at two o'clock in the morning while in a highly intoxicated condition. ☐

TEST XXIX

Below are listed several crimes that have been committed at various times. Disregarding any legal consequences, mark with a "1" that which you condemn the least and with a "5" the one that you condemn the most and rate the rest accordingly.

- a. A man in a high position in public life accepted bribes for the dispensation of government favors which resulted in a loss of millions to the government. ☐
- b. A man with the help of some accomplices broke into the home of a hermit and tortured him with knives and hot irons until he died, in an attempt to make him tell where he had hidden his money. ☐
- c. A contractor, furnishing materials to the government, was able, by means of short weights and inferior qualities, to defraud the government out of a million dollars. ☐
- d. A politician started a malicious story about a senator which had no foundation in fact but which was so generally believed that the senator failed of re-election. ☐
- e. A high government official divulged certain military secrets to a rival government and netted himself a half-million dollars in the transaction. ☐

TEST XXX

Following are some activities engaged in by some boys of high school age. Disregarding the legal consequences, mark with a "1" the activity which you would condemn the least and with a "5" the activity which you would condemn the most and rate the rest accordingly.

- a. Stole a large car and drove it around until it was out of gasoline and then left it on the street. ☐
- b. Set fire to a vacant house so that they could watch the fire department put out the flames. ☐
- c. Threw shingle nails by the handful on a boulevard so that they could watch the discomfort of the motorists. ☐
- d. Bet all the money they could on the football team of the opposing school and agreed that although they were supposed to play on their own team, they would let the other school win so that they could win their bets. ☐
- e. Driving a large car on a lonely road, they collided and badly wrecked ☐

a small car containing several people and then sped away from the scene without giving aid. ☐

TEST XXXI

Below are five of the ten commandments as given in the new American version of the Bible. Disregarding any legal consequences or anything that you have learned in church or school, put a "1" after the one that you think is the least important and a "5" after the one that you think is the most important and rate the rest accordingly.

- a. Honor thy father and mother that you may live long in the land the Lord your God is giving you. ☐
- b. You must not commit murder. ☐
- c. You must no commit adultery. ☐
- d. You must not steal. ☐
- e. You must not bring false charge against your fellow. ☐

REFERENCES

1. ALLPORT, G. W. Personality and character. *Psychol. Bull.*, 1921, **18**, 441-445.
2. BROGAN, A. P. A study in statistical ethics. *Int. J. Ethics*, 1923, **33**, 119-134.
3. CADY, V. M. Psychology and pathology of personality. *J. Delinq.*, 1922, **7**, 225-248.
4. CLARK, W. W. The measurement of social attitudes. *J. Appl. Sociol.*, 1922, **10**, 13-20.
5. FERNALD, G. G. The defective delinquent class: differentiating tests. *Amer. J. Insan.*, 1912, **68**, 523-594.
6. HARTSHORNE, H., & MAY, M. A. Personality and character tests. *Psychol. Bull.*, 1926, **23**, 395-411.
7. HARTSHORNE, H., MAY, M. A., & WELTY, R. E. Personality and character tests. *Psychol. Bull.*, 1927, **24**, 418-435.
8. LASK, L. J. Ueber die Fernald'sche Methods zur Prüfung des sittlichen fühlens. *Beihefte Z. Zsch. f. angew. Psychol.*, No. 24. Pp. 1-84.
9. RAUBENHEIMER, A. S. An experimental study of some behavior traits of the potentially delinquent boy. *Psychol. Monog.*, 1925, **34**, No. 159. Pp. 107.
10. THURSTONE, L. L. The method of paired comparisons for social values. *J. Abn. & Soc. Psychol.*, 1926, **21**, 384-400.

Washington University
St. Louis, Missouri

UN TEST COLLECTIF DE CRUAUTÉ-COMPASSION POUR LA DÉCOUVERTE DES CRIMINELS POTENTIELS

(Résumé)

Cette étude a eu comme but la formulation d'un test qui déterminerait les tendances qui prédisposeraient un individu vers les crimes du type contre-la-personne. Bien que l'analyse rationnelle montre qu'il existe plusieurs facteurs prédisposants qui pourraient rendre un individu capable de commettre ce type de crime, tous ces facteurs semblent une partie de ce qu'on pourrait nommer un continuum cruauté-compassion. Pour trouver une mesure de la qualité de cruauté-compassion d'un sujet donné, on a formulé un test composé de trente-et-un groupes, chacun de cinq parties. On a demandé aux sujets d'évaluer chacun de ces groupes en ordre de préférence de 1 à 5. Chaque groupe contenait une partie distinctement sadiste. La somme des évaluations données à cette partie sadiste est le résultat du test. On a standardisé le test avec des sujets de la "high school" ordinaire et on a comparé les résultats avec ceux des sujets des Écoles de Correction. On a obtenu des différences aussi grandes que 9 fois l'Erreur Probable de la différence. Les résultats peu élevés (indiquant des tendances cruelles) ont été comparés aux histoires des cas qui ont montré de l'évidence de comportement criminel évident dans presque tous les cas. On offre donc le test comme moyen d'évaluer les individus sur un continuum cruauté-compassion, et les résultats semblent montrer que ceux à l'extrémité cruauté de l'échelle sont plus prédisposés à commettre les crimes contre-la-personne. Il est à espérer qu'au moyen d'un tel outil on peut mettre à part les individus disposés à la cruauté et leur donner un traitement hygiénique avant qu'ils ne commettent des crimes.

HAWTHORNE

EIN GRUPPENTEST FÜR GRAUSAMKEITSMITGEFÜHL FÜR DIE ERMITTLUNG VON POTENTIELLEN VERBRECHERN

(Referat)

Es war der Zweck dieser Untersuchung, einen Test zu erfinden, der die Neigungen bestimmen könnte, die ein Individuum auf Verbrechen gegen die Person vorbereiten. Obschon die rationelle Analyse zeigt, dass es eine Reihe von vorbereitenden Faktoren gibt, die ein Individuum befähigen könnten, Verbrechen dieser Art zu begehen, können doch alle diese Faktoren auf das zurückgeführt werden, was als kontinuierliches Grausamkeitsmitgefühl (cruelty-compassion continuum) bezeichnet werden möchte. Um die Grösse des Grausamkeitsmitgefühls irgend eines gegebenen Individuums zu messen, wurde ein Test erfunden, der aus einunddreissig Gruppen, und jede aus fünf Punkten bestand. Die Versuchspersonen wurden aufgefordert jede dieser Gruppen nach der Ordnung ihrer Wahl von 1 bis 5 zu schätzen. Jede Gruppe enthielt einen Punkt, der einen deutlichen sadistischen Einschlag hatte. Die Summe der Schätzungen, die dem sadistischen Punkt beigelegt wurden, ist der Testkoeffizient. Der Test wurde an Versuchspersonen der regulären Höheren Schulen normiert, und die Koeffizienten verglichen mit denjenigen der Versuchspersonen von Korrekptionsanstalten. Es wurden Unterschiede festgestellt, die neun mal die Fehlerquelle des Unterschiedes betrug. Niedrige Koeffizienten (die grausame Neigungen anzeigen) wurden gegen die Geschichten einzelner Fälle abgewogen, die in fast allen Fällen

schon Beweise offenkundiger verbrecherischer Haltung anzeigten. Der Test wird darum als Mittel dargeboten, um Individuen auf das kontinuierliche Grausamkeitsmitgefühl zu prüfen. Die Ergebnisse scheinen anzuzeigen, dass diejenigen, die am Grausamkeitsende der Skala stehen, mehr dazu veranlagt sind, Verbrechen gegen die Person zu verüben. Man spricht die Hoffnung aus, dass durch ein solches Mittel die grausam veranlagten Individuen abgesondert, und vor dem Begehen offener Verbrechen hygienisch behandelt werden können.

HAWTHORNE

NERVOUS INSTABILITY AND GENIUS: MILITARY AND POLITICAL LEADERS*

*From the Psychological Laboratories of Northwestern University
and Ohio University*

PAUL A. WITTY AND HARVEY C. LEHMAN

Historical accounts frequently become exaggerated or distorted with the passing of time. It is natural for everyone to conceal or underestimate his weaknesses; therefore, autobiographies are frequently unreliable guides in an evaluation of the conduct and character of leaders. Biographies often portray only the most obvious features in the lives of men; consequently, he who describes with great accuracy the motivation of behavior is indeed astute. Biographies concerning Jonathan Swift exemplify clearly the limitations of attempts to identify the motivating agencies which determined the conduct of this man. Volumes have been written regarding Swift. And still he remains an enigma.

That Swift was the most powerful leader of his time is sometimes stated; he kept his affiliation with the church, and, nevertheless, he became the real ruler of Ireland. Swift was an alien during his entire life, a misanthrope who conceived his amazing couplets to harass and annoy his contemporaries. He appeared to write, not for the joy of creation, nor with literary values in mind, but to satisfy animosities. Accuracy was, of course, not considered in many of his attacks, but his astute mind rightly viewed and keenly analyzed many of the most pressing problems of his day. His classic *Gulliver's Travels* is a diatribe, resulting from his contemptuous attitude toward his contemporary fumbler in political and social affairs. That this book resulted largely from animosities and maladjustment cannot be doubted.

The exact cause of Swift's maladjustment is unknown. Thwarting of sex desires was doubtless one element; disease was perhaps another; failure to achieve recognition in early life another—an interminable list of causal factors might be assembled. Whatever were the true

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, March 30, 1931.

sources of his maladjustment, the fact stands out that this man developed classic literature in order to compensate or to atone for his denials. Carl Van Doren's (18) excellent and somewhat brilliant biography leaves many of the baffling questions unanswered. Van Doren is unusually careful in his analysis. The psychoanalyst, unlike Van Doren, often takes the frailest fragments and weaves a fascinating yet unreliable account of the motivation of conduct. It is therefore hazardous to place too much credence in the various biographies and narratives concerning the private lives and human frailties of the great and the near great.

We have previously heard that the eccentricity of genius is largely a myth (16, p. 184). Some think that this statement has been substantiated by Miss Cox's analysis of biographical accounts of acknowledged geniuses (6). Nevertheless, there is evidence of the existence of eccentricity in certain eminent military and political leaders, and there is a growing tendency among psychoanalysts and certain historians to ascribe a part at least of the achievement of these men to eccentricity or abnormality. It is of interest that, although the psychoanalyst directs his attention to the eccentricities of the great or near great, certain psychologists are loudly proclaiming that eccentricity and genius are not correlated. Indeed, some are certain that the genetic studies of gifted children have dispelled entirely the early theories regarding the relation of instability and genius. The following quotation exemplifies the extreme to which this attitude may extend.

"The publication of this volume [*Terman's Genetic Studies of Genius. Vol. I.*] should lay forever at rest the lingering ghosts of the old theories and superstitions regarding the disadvantages and dangers of precocity, and the relation of genius to nervousness, moral or emotional instability, mental aberration or physical defect" (8, p. 174).

The writers of this paper previously have suggested that the thwarting of certain strong desires often results in a redirection of energy and sometimes in conspicuous attainment (20). The thwarting of desires is, of course, only one of numerous elements which elicit drive or which intensify effort toward particular goals. The writers have pointed out that inborn and organismic urges actuate drive for some; that habits, once acquired, become the impetus for drive in the cases of others, and that the mal-functioning of glands may be responsible in some persons for intensified effort (19). The writers have asserted

also that, in numerous instances of so-called greatness (of certain types), there are unmistakable evidences of the desire to compensate for thwarted wants (19). Some thwarted individuals find satisfyingness in religious frenzy. Others turn to music for substitute gratification, and some find vicarious satisfaction in creating imaginative products in which the sublimated wish is to varying degrees fulfilled (21, 22). It is apparent that the forms of substitute activity are numerous and varied.

Among the world's great military leaders and statesmen there are many in whom the early thwarting of natural wants and desires appears to have been an important factor in determining not only drive, but also the nature of the goal. In some instances the inability to endure the youthful thwarting probably resulted in unusual and persistent drive. This drive, combined with a reasonable amount of ability and opportunity, sometimes enabled these individuals to attain the rank of recognized genius.

MELANCHOLIC LEADERS

Cromwell's life is characterized by two very important qualities. Cromwell was intensely religious, and he was relentless, merciless, and cruel. These characteristics make him appear paradoxical. They were, perhaps, fortunate correlates in this man, for he could punish and torture his fellowmen mercilessly, and then compensate rather completely by means of religious fervor. Nevertheless, even as a boy, Cromwell felt the limitations of his religious faith. He struggled with his conflicting impulses and beliefs, and finally developed a satisfactory belief. For some years he was harassed by religious conflict, from which he emerged in a triumphant conversion. For the remainder of his life he was deeply religious.

As an adult, Cromwell was an unusually successful military leader. Moreover, he was exceedingly merciless. At Groheda, he ordered the massacre of 2800 men who had refused to capitulate. In his treatment of Ireland, also, he was pitiless. Mental conflict followed, of course, his extreme cruelty. He therefore revealed many inconsistencies. He became progressively more religious; and yet he remained incredibly cruel. Therefore, it was inevitable that mental conflict would occur repeatedly. Mental conflict, of course, gives rise to many types of compensatory adjustments. Religion appeared to have been the most adequate method for reconciling Cromwell's qualms. This reconciliation, however, was not entirely successful,

and Cromwell developed and was dominated by a number of delusions. One of these was his belief that the Catholic powers might combine and crush Protestantism; Cromwell was constantly negotiating with Charles X of Sweden, whom he erroneously regarded as a second Gustavus Adolphus, to form a counter-Protestant league.

That Cromwell himself was afflicted with symptoms of instability is clear. Thomas Carlyle gave the following account of him.

"What little we know of his earlier obscure years, distorted as it has come down to us, does it not all betoken an earnest, affectionate, sincere kind of man? His nervous melancholic temperament indicates rather a seriousness *too* deep for him. Of these stories of 'Spectres'; of the white Spectre in broad daylight, predicting that he should be King of England, we are not bound to believe much;—probably no more than of the other black Spectre, of the Devil in person, to whom the Officer saw him sell himself before Worcester Fight. But the mournful, over-sensitive, hypochondriac humour of Oliver, in his young years, is otherwise indisputably known. The Huntington Physician told Sir Philip Warwich himself, He had often been sent for at Midnight; Mr. Cromwell was full of hypochondria, thought himself near dying, and 'had fancies about the Town-cross.' These things are significant. Such an excitable deep-feeling nature, in that rugged stubborn strength of his, is not the symptom of falsehood; it is the symptom and promise of quite other than falsehood" (2, p. 294).

Undoubtedly, Cromwell was nervously unstable in his childhood. The instability appears to have been a conspicuous feature of his entire life. Compensations, therefore, of various sorts were demanded continuously; they were, however, required much more frequently and urgently in adult life, when mental conflicts were numerous and inevitable. Thus, Cromwell became a relentless leader who tried to atone for his mercilessness by religious devotion. The adjustment appeared unsuccessful and one finds Cromwell, at one time, an unrelenting persecutor, at another, a pitiful melancholic creature.

The melancholic type of military leader is frequently encountered in the biographical literature. Of course, biography is often unreliable, and slight irregularities in conduct are exaggerated. Many times, however, the biographer has an extremely sympathetic attitude toward his subject; he is likely, in such a case, to minimize or discount tendencies which play an important rôle in determining the career. It is, of course, difficult to assign the label of genius to

leaders of war. These leaders often are products of momentary need, of driving power, and of group hysteria; they are therefore difficult to evaluate intellectually. The writers feel that many of the military leaders have been products of capricious circumstances; that the endeavor of many of these men has been simply the bungling of situations which chance turned to success. Many military leaders were men whose intellectual calibre certainly could not be esteemed highly from their blundering, blustering careers. The late war for democracy (with its resultant world fervor) was characterized by wasteful, unintelligent leadership. A war makes leaders, but it is extremely doubtful that the leader is chosen from the ranks of the intelligent. Two types of military leaders are conspicuous: the blundering, dynamic bully, and the sober, melancholic stoic. The first is ubiquitous in biographies of war heroes. Lincoln is a striking example of the second type.

LINCOLN

The life and behavior of Abraham Lincoln are difficult indeed to analyze and evaluate. It is quite generally agreed, however, that Lincoln was characterized by extraordinary melancholy.

"He was a sad-looking man; his melancholy dripped from him as he walked. His apparent gloom impressed his friends,¹ and created sympathy for him—one means of his great success" (9, Vol. II, p. 297).

The foregoing description was written by Herndon prior to psychoanalysis. Clark has sought to give a psychoanalytic interpretation to Lincoln's life. He attributes Lincoln's difficulties to a life-long "mother fixation" (5). There is some evidence to support this hypothesis. There was but one instance of a really sentimental attachment in the life of Lincoln until he was 25 years of age; then he

¹"Lincoln's melancholy never failed to impress any man who ever saw or knew him. The perpetual look of sadness was his most prominent feature. The cause of this peculiar condition was a matter of frequent discussion among his friends. . . The reader can hardly realize the extent of his peculiar tendency to gloom. One of Lincoln's colleagues in the Legislature of Illinois is authority for the statement coming from Lincoln himself that this 'mental depression' became so intense at times that he never dared to carry a pocket knife.' Two things greatly intensified his characteristic sadness; one was the endless succession of troubles in his domestic life, which he had to bear in silence; and the other was unquestionably the knowledge of his own obscure and lowly origin. The recollection of these things burned a deep impress on his sensitive soul" (9, Vol. II, p. 297, footnote).

met Anne Rutledge. At this time, she was engaged to another. In view of Lincoln's knowledge of this, it seemed to have been relatively easy for him to form an attachment for Miss Rutledge, easier than if she had been entirely free and marriageable.

The affair with Miss Rutledge ended unfortunately with her early death. Herndon has described the effect upon Lincoln.

"The most astonishing and sad sequel to this courtship [with Anne Rutledge] was the disastrous effect of Miss Rutledge's death on Mr. Lincoln's mind. It operated strangely on one of his calm and stoical make-up. As he returned from the visit to the bedside of Miss Rutledge, he stopped at the house of a friend, who relates that his face showed signs of no little mental agony. 'He was very much distressed,' is the language of this friend, 'and I was not surprised when it was rumored subsequently that his reason was in danger.' One of Miss Rutledge's brothers says: 'The effect upon Mr. Lincoln's mind was terrible. He became plunged in despair, and many of his friends feared that reason would desert her throne. His extraordinary emotions were regarded as strong evidence of the existence of the tenderest relations between himself and the deceased.' The truth is that Mr. Lincoln was strangely wrought up over the sad ending of the affair. He had fits of great mental depression, and wandered up and down the river into the woods woefully abstracted—at times in the deepest distress. If, when we read what the many credible persons who knew him at the time tell us, we do not conclude that he was deranged, we must admit that he walked on that sharp and narrow line which divides sanity from insanity" (9, Vol. I, p. 129).

At the age of 31, Lincoln became engaged to Mary Todd, but his inability to overcome his earlier denials and melancholia was apparent. He failed to appear at the wedding; his friends, after extended search, finally discovered a miserable fearful creature, whose mental state bordered upon insanity.

"By daybreak, after persistent search, Lincoln's friends found him. Restless, gloomy, miserable, he seemed an object of pity. His friends, Speed among the number, fearing a tragic termination, watched him closely day and night. Every instrument that could be used for self-destruction was removed from his reach. Mrs. Edwards did not hesitate to regard him as insane, and her sister, Miss Todd, shared in that view" (5, p. 12).

"Here one is at once struck with the fact that the depression immediately succeeding this episode was one in which there was

not only an incomplete adjustment to Miss Todd as a bride, but, as we shall see later, Lincoln seemed unable to adapt himself to the full requirements of marriage itself. Hence the profound depression which came on at this time. This despondency was as deep as that ordinarily seen in the depressive psychosis; there was retardation in thought and action, periods of extreme silence, listlessness, indifference, loss of appetite, insomnia, alternating with moods of anxious restlessness. He had gloomy forebodings and thoughts of suicide. This depressive period extended over nearly ten months following his failure to appear on the date first set for his marriage. During this time Lincoln was absent from his regular duties in the State assembly, which he had up to this time carefully and painstakingly attended; he was finally taken by his good friend, Speed, to Kentucky to regain his health. The question has often been asked whether Lincoln actually had a short period of detention in a sanitarium. We have information, however, that he was kept under the careful watch of different members of the family in Kentucky and was permitted to occupy himself as he pleased upon the ranch" (5, pp. 12f.).

Clark concludes his study of Lincoln by stating that Lincoln suffered from an infantile fixation for his mother. The evidence in support of this is certainly impressive even though it be inconclusive.

"From this study, then, it would seem that no small part of Lincoln's depression was due to certain deep, unconscious fixations of soul attachment to the mother hindering the normal emotional life which in turn made it impossible in early life for him to assume the usual attitude of religious feeling and thought" (5, p. 18).

Clark's study, which is to some degree corroborated by others, reveals unmistakably the life-long conflict which characterized the man. The unusual behavior in childhood was followed by aberrations in conduct during later years. Whether or not one attributes the unusual conduct to a mother fixation, one is forced to realize that nervous instability and sex maladjustment were significant factors in determining Lincoln's activity throughout his entire life. His impelling drives are definitely traceable to youthful thwarting and attendant anxieties. These anxieties characterized Lincoln always, and undoubtedly played an important rôle in determining the direction of his energy and the nature of his accomplishment.

JULIUS CAESAR

Julius Caesar's renown as a military leader is unquestioned. No list of military geniuses would be complete unless it included Caesar's name. Does it follow that Caesar was a stable and well-balanced individual? Numerous facts seem to preclude the acceptance of this hypothesis.

It is frequently alleged that Caesar manifested homosexuality. This charge probably is not particularly significant, because in Caesar's day it was quite common for boys of social status to have strong attachments for each other. It is said that Caesar had numerous love affairs with members of both sexes. All of this is of interest, but from the standpoint of eccentricity it is not particularly significant.

More startling is the following statement of Suetonius. Suetonius records a dream of Caesar which doubtless suggests instability. Probably this dream would not have been recorded but for the ancient custom of hiring interpreters to reveal the meanings of dreams.

"In the stillness of the night following, he dreamt that he lay with his mother; but his confusion was relieved, and his hopes were raised to the highest pitch, by the interpreters of his dream, who expounded it as an omen that he should possess universal empire; for that the mother who in his sleep he had found submissive to his embraces, was no other than the earth, the common parent of all mankind" (14, pp. 5 f.).

Certainly such a dream was indicative of something other than normal sex interest. Regardless of the proper psychoanalytic interpretation, the writers insist that recollection of and anxiety regarding such an extraordinary dream does connote considerable eccentricity.

Caesar is said also to have suffered from some sort of epilepsy. One of his recent biographers states that, even in childhood, he was subject to slight attacks (17, pp. 59 f.). His physicians diagnosed the trouble as a mild distemper which would be outgrown. Nevertheless, Caesar's nervous affliction increased as he matured. Several accounts record the increasing frequency of attacks with increasing maturity. At Rhodes, Caesar fell several times to the ground, and he displayed decidedly marked epileptic symptoms (17, pp. 59 f.).

Within Caesar's immediate family there was a difference of opinion as to the cause of his strange behavior. His mother and his sister insisted that his illness was the result of dissipation. Caesar pre-

ferred, however, to believe that he had inherited his weakness from his father, since the father had experienced similar attacks.

In spite of the fact that Caesar felt that his illness was a constitutional weakness, he admitted that the affliction might have been aggravated by riotous living. He therefore adopted a frugal diet and devoted himself to strenuous outdoor exercise.

Caesar became a relentless worker. Plutarch expresses the belief that his tendency to be incessantly active was merely an attempt to protect his body from further attacks of some form of congenital weakness. Plutarch probably would have insisted that Caesar accomplished what he did, not in spite of his weakness, but because of it.

" . . . he was . . . distempered in the head and subject to an epilepsy, which, it is said, first seized him at Corduba. But he did not make the weakness of his constitution a pretext for his ease, but rather used it as the best physic against his indispositions; whilst, by indefatigable journeys, coarse diet, frequent lodging in the field, and continual laborious exercise, he struggled with his diseases, and fortified his body against all attacks. He slept generally in his chariots or litters, employing even his rest in pursuit of action. In the day he was thus carried to the forts, garrisons, and camps, one servant sitting with him, who used to write down what he dictated as he went, and a soldier attending him with his sword drawn" (7, pp. 561 f.).

Certain present-day psychiatrists and physicians assert that epilepsy speeds metabolism and develops drive. If this be valid, it may be that Caesar's epilepsy was an important factor in determining the extent and the nature of his accomplishment.

Caesar's epilepsy probably accounts in part, not only for his incessant activity, but also for his unsympathetic and cruel treatment of others. In describing Caesar's activities in Gaul, Thaddeus writes:

"His eighth year is spent suppressing independent uprisings. His ferocity is now at times almost that of a madman. His epileptic fits, from which he had been free for several years, have returned with greater ferocity than ever before. When he broods over the thought that the unrest in Gaul may cheat him of his ambition to be king his fury against the Gauls is ungovernable. . . With bloody strides he crosses and recrosses Gaul depopulating it of those who refuse to submit to his will" (17, p. 216).

The evidence that Caesar suffered from some unusual "falling sickness" is of several kinds. Cleopatra sought to assure Caesar that his epileptic attacks were not ordinary mortal sickness, but that they were divine seizures (17, pp. 247 f.). And Plutarch records that, at one time, when the Roman senate conferred upon Caesar certain extravagant honors, he did not rise. This act offended greatly the consuls and the praetors who had waited upon him. As soon as Caesar realized that he had committed a serious offense, he left for his home. He subsequently excused himself for not rising, explaining that those who are attacked by the malady from which he was suffering lose their presence of mind if they talk much while standing; that they also soon grow giddy under such circumstances and fall into convulsions, quite losing their reason (7, pp. 593 ff.).

Whether Caesar had ordinary *grand mal* epilepsy or, as suggested by MacLaurin, *petit mal*, it seems reasonable to infer that his nervous instability was a causal factor of no small import in determining his military success.

FREDERICK THE GREAT

Frederick II was also a tireless worker. When Frederick was king of Prussia, occasional conversation and musical endeavor were his only relaxations. His excessive activity was characteristic of him even in youth. Frederick wrote well at an early age. His father attempted to repress his writing, and a struggle between father and son ensued. Frederick II at one time made an unsuccessful attempt to flee to England; the circumstances reveal smouldering hatred and resentment toward his father. Frederick's accomplice was beheaded in his presence, and Frederick was led to believe that he also was to be beheaded. Although the father did not carry out this threat, he did place Frederick under exceedingly strict régime. And Frederick was forced subsequently to marry against his will. All resulted in life-long hatred of the father.

As king of Prussia, Frederick's behavior certainly does not tend to substantiate the hypothesis that the eccentricity of genius is a myth. As a military leader, Frederick must certainly be classified as a genius. MacLaurin draws upon one of Frederick's most recent biographers for a pointed description.

"In later years Frederick gave up shaving, and merely clipped at his beard with scissors. He seldom washed any part of his person, even his head and face. In this respect he was

very different from his father,² who used soap and water freely, and often complained of his son's dirtiness. One of his valets concluded from his master's dislike of water that he must be afflicted with a kind of hydrophobia. . . His complexion, was tanned. . . doubtless because it was seldom washed, like a tramp's today; but unlike a tramp's, it was touched up with red paint' " (11, pp. 194-195).

MacLaurin asserts that at the time of his death Frederick was so dirty that it was difficult to get his body ready for the grave. Frederick tried to avoid every habit of his father. His life habits and military successes probably were protests of his resentful, unstable nature. Most boys probably would have conformed to parental desires if they had been in Frederick's place. But Frederick deviated in his behavior; his instability and resultant deviation in conduct combined to produce another military leader.

NAPOLÉON

Napoleon is, without doubt, one of the world's most brilliant military leaders. In the matter of slaying his fellowmen he appears to have reached the summit of human achievement. He caused the deaths of approximately 3,000,000 of his own men and millions of his enemies. He fought in 60 battles, of which he lost only two! When he was at the peak of his power he controlled more than half of Europe. For more than twenty years the whole civilized world followed his career in fear and awe. His colossal achievements as a soldier have given rise to a hero-worship that is almost unparalleled. Surely, Napoleon's ability in all matters of military organization and administration entitles him to rank as a foremost military genius. Intelligence alone can scarcely account for the activities of this powerful world figure. Napoleon stood forty-second in his class in the military academy. Moreover, he sometimes revealed a naïvete in his public utterances.

Drive alone might have sent Napoleon to the gallows or to an early defeat. Opportunity singly could not have effected his rise to power, for certain other men of his time must have had as great an opportunity as he. There seems to have existed a peculiar concatenation of factors which caused Napoleon to become a world figure. Undoubtedly, he was characterized as much by his drive as by any other

²A Freudian might say that Frederick's dirtiness simply represented an unconscious revolt against the tyranny of his father.

quality. This statement is not to be interpreted as a denial of Napoleon's great military ability. It is made to concentrate attention for the moment upon an element which was certainly no less important than latent capacity in determining the activity of this man.

Napoleon is said to have asserted that genius is industry, meaning, of course, that genius is industry plus other things. Many witnesses have commented upon Napoleon's amazing power for continuous and arduous work. Ludwig writes dramatically of him.

"He robbed hundreds of his fellow workers of health and youth, because he demanded too much of them when he demanded from them what he exacted from himself. His private secretary would be sent for at a late hour, and would get to bed at four in the morning; at seven, the poor man would find new tasks ready for him, and would be told that they must be finished within two hours. When Napoleon and his secretary were together all day, one dictating and the other writing from dictation, at meal times the chief would order food for two and would share with his subordinate at a corner of the work-table, just as he would have shared with his adjutant on a boundary stone. During the Consulate he would sometimes begin a sitting with his ministers at six in the evening and keep it up till five next morning. In the three months at Schonbrunn, his official correspondence comprised four hundred and thirty-five letters occupying four hundred folio pages of print. This was only his political and administrative correspondence; in addition he wrote a great many private letters, and delivered innumerable orders by word of mouth" (10, p. 573).

Napoleon once described his habits of work frankly:

"I am always at work; I think a great deal. If I appear to be ever ready and equal to any occasion, it is because I have thought over matters long before I undertake to do the slightest thing: I have foreseen eventualities. There exists no guardian angel who suddenly and mysteriously whispers in my ear what I have to do or say. Everything is turned over in my mind, again and again, always, whether I am at table or at the theatre. At night, I wake up in order to work" (10, p. 569).

Not only did Napoleon work continuously; he also worked rapidly. The king of Prussia has depicted this tendency: "We need but see him ride: he always gives his horse free rein, and never troubles about what may be happening in his rear" (10, p. 571). Napoleon's mental

work was done with unusual speed. At St. Helena he once remarked that, as emperor, he was accustomed to dictate on different topics to four secretaries at once (10, p. 656). As a youth, Napoleon also was a tireless worker. At age sixteen, he wrote:

"I have no refuge but my work. I only change my linen once a week. Since I was ill, I have slept very little. . . . I eat only once a day" (10, pp. 12 f.)

In the foregoing quotation Napoleon speaks of work as a refuge. A refuge from what? Napoleon was noted for his indifference to danger upon the battle field. Why, then, should this military leader have sought refuge? Examinations of his early years throws light upon this interesting question.

The student of history will recall that when Napoleon was nine years of age he was admitted to the school at Brienne as a sort of charity pupil. His parents, although they belonged to the Corsican nobility, were extremely poor. Napoleon's education was therefore financed by his French protector, Count de Marboeuf. At the age of nine Napoleon thus found himself a stranger in a strange land, among a people whom he detested, because they were the oppressors of his countrymen. Another difficulty disturbed him. He spoke French poorly. His new companions jeered at him because of his pronunciation. His name was strange; *la paille au nez* was the nickname they made for Napoleon. And the contemptuous treatment he received because of his poverty was such that he begged to be taken home. He wrote as follows to his father (15, pp. 22 ff.).

"My father, if you or my protector cannot give me the means of sustaining myself more honorably in the house where I am, please let me return home as soon as possible. I am tired of poverty and of the jeers of insolent scholars who are superior to me only in their fortune, for there is not one among them who feels one hundredth part of the noble sentiment which animates me. Must your son sir, continually be the butt of these boobies, who, vain of the luxuries which they enjoy, insult me by their laughter at the privations which I am forced to endure? No, father, no! If fortune refuses to smile upon me, take me from Brienne, and make me, if you will, a mechanic. From these words you may judge of my despair. This letter, sir, please believe, is not dictated by a vain desire to enjoy extravagant amusements. I have no such wish. I feel simply that it is necessary to show my companions that I can procure them as well as they, if I wish to do so.

"Your respectful and affectionate son,
"BONAPARTE."

Young Napoleon found his position among the French boys intolerable. But the foregoing letter brought him no relief. It was necessary, therefore, for him to find another means of escape. Even more unbearable than his poverty were the taunts that he suffered as a member of a subject race. His comrades were, of course, French, and it was the French who had conquered Corsica. Once when they harassed him by stating this fact, he told them that had there been four to one, Corsica would never have been conquered, but that the French came ten to one. When they repeated the taunt, "But your father submitted," young Napoleon replied bitterly, "I shall never forgive him for it" (15, pp. 21f.).

The boys on one occasion brought the sweeping charge of cowardice against all the inhabitants of Corsica in order to exasperate him. Sloane has described Napoleon's unhappy plight in the following:—

"The climax of the miserable business was reached when to a taunt that his ancestry was nothing, his father a wretched tipstaff, Napoleon replied by challenging his tormenter to fight a duel. For this offense he was put in confinement while the instigator went unpunished" (13, p. 52).

Upon being imprisoned, Napoleon wrote immediately to his protector begging him to withdraw his financial support. Napoleon knew that such withdrawal would result in his being sent home. Marboeuf, however, did not accede, and Napoleon found that he must remain at Brienne.

Some persons, if placed in Napoleon's position, would have reacted by admitting quietly actual inferiority. Not so with Napoleon! It must be remembered that Napoleon was proud and sensitive, and that he was quite incapable of withstanding continuous thwarting. In his later life he once remarked: "I am a man whom people may kill but will not affront" (10, p. 556). Here, then, was an extremely sensitive individual who as a lad was humiliated by the following facts:

He could not speak well the language of his schoolmates.

He belonged to a subject race.

His own father had had a share in uniting Corsica to France. This the boy resented.

His father occupied a station in life more humble than those of the fathers of his schoolmates.

He was extremely poor and the fact was plainly evident to his schoolmates.

He was openly taunted by his companions who held themselves to be superior.

He was unsuccessful in his school work.

Is it any wonder that defense reactions were set up? Ludwig describes Napoleon's student days as follows:—"He is taciturn, with a love for solitude; is moody, over-bearing, and extremely egotistical" (10, p. 9). The official report represents him as silent and obstinate. Small wonder that this proud youth reacted with sullen, morose behavior! Was this not a defensive pose, and one which today is easily understood? Sloane describes Napoleon's existence at Brienne.

"Dark, solitary, and untamed, the new scholar assumed the indifference of wounded vanity, despised all pastimes, and found delight either in books or in scornful exasperation of his comrades when compelled to associate with them. There were bitter fights, in which the Ishmaelite's hand was against every other. Sometimes in a kind of frenzy he inflicted serious wounds on his fellow-students. At length even the teachers mocked him, and deprived him of his position as captain in the school battalion" (13, p. 52).

The foregoing picture is corroborated by practically every biographer, as is also the following.

"Each of the hundred and fifty pupils had a small garden spot assigned to him. Bonaparte developed a passion for his own, and, annexing by force the neglected plots of his two neighbors, created for himself a retreat, the solitude of which was insured by a thick and lofty hedge planted about it. To this citadel, the sanctity of which he protected with a fury half insane, he was wont to retire in the fair weather of all seasons, with whatever books he could procure" (13, pp. 52 f.).

Here, then, we see the sort of refuge sought by the youthful Napoleon; a refuge safe from the gibes and the insults of his schoolmates. Sloane records that, in the companionship of his books, Napoleon spent happy, pleasant, and fruitful hours. He became a wide reader and an assiduous mental worker. And these habits continued in later years. As a young lieutenant, he asserted that he had no refuge but his work. Later he stated that work was his element; that he had been created for it. Few men have worked as hard as did Napoleon. Roederer, who was his close companion during the Consulate, wrote:

"That which especially characterizes him is the power and persistence of his attention. He can work for eighteen hours at a stretch, it may be at one piece of work, it may be at several in turn. I have never seen his mind flag. I have never seen his mind without a spring in it, not when he was physically tired, not when he was taking violent exercise, not even when he was angry. I have never seen him distracted from one affair by another, neglecting the matter in hand for one which he is about to work. Good or bad news from Egypt never interfered with the steps it was necessary to take for the safety of Egypt. No one was ever more wholly immersed in what he was doing, nor did any one ever make a better distribution of his time among all the things he had to do. Never was anyone more stubborn in rejecting the occupation, or the thought which was not appropriate to the hour of the day; nor was anyone ever more adroit in seizing an occupation or a thought when the right moment had come'" (10, pp. 572-573).

Napoleon's tendency to read, study, and work probably received early impetus as a result of the mortification received at the hands of his more fortunate classmates at Brienne. There is evidence that his whole outlook upon life was colored by these early experiences. For example, in later life, Napoleon professed contempt for legitimate heirs to thrones. He asserted that those who inherit thrones are usually soft drones and far inferior to the men who win crowns by their own strength. At the same time he sought to establish all of his relatives among the nobility; this was probably one of the results of a very deep-seated desire for family recognition. He refused, however, to take an active interest in his own genealogy. He felt that he was more important than any of his ancestors; he felt insulted when anyone questioned his ancestry. As a youth, ancestry had been to him a painful subject. In spite, however, of his contempt for the legitimate heir to a throne, he was most anxious that his own son might succeed him. Napoleon was greatly troubled by the problem of establishing a dynasty. When he was at the height of his power, this was one of his chief problems.

Thus we see in Napoleon some strange inconsistencies. These are understandable in part in the light of his boyhood experiences. At Brienne he made few friends. To one of these comrades, Bourrienne, he was attached for years. He remarked once to Bourrienne, "You never laugh at me; you like me."

Those who found Napoleon morose and surly as a boy did not realize that beneath the sullen exterior there was a proud nature longing for appreciation and demanding recognition; that it was sensitiveness rather than arrogance which drove him away from his schoolmates. Always Napoleon sought and lived for recognition and adoration. As emperor, Napoleon still craved homage from the Brienne group. At St. Helena he remembered how his German teacher at Brienne had treated him with contempt. He remarked, "I should very much like to know whether Herr Bauer ever learned how I made good" (10, p. 556).

This discussion suggests that the early thwarting had a marked and continuous influence upon Napoleon's conduct and thought in later life. There were, of course, additional factors that influenced him greatly. Ludwig, MacLaurin, and others state that Napoleon was abnormal sexually. Ludwig writes:

"The doctors have told us a good deal about his physical condition. 'Pulse never more frequent than 62; bosom well padded, almost like a woman's, and with very little hair; *partes viriles exiguities insignis sicut pueri*'" (10, p. 547).

And MacLaurin states:

"Napoleon's hands and feet were extremely small; his skin was white and delicate; his body had feminine characteristics, such as wide hips and narrow shoulders; his reproductive organs were small and apparently atrophied. He is said to have been impotent for some time before he died" (12, pp. 215.)

The foregoing assertions fit in well with the facts. It is well known that Napoleon took comparatively little advantage of his opportunities for sexual relations with women. He was fond of saying: "Love is not for me, I am not like other men" (1, p. 10). Between the Cairo days and the time of the affair with the Countess Walewski, Napoleon never had a mistress, but his generals had many (10, p. 264). In Egypt he once spoke to one of his confidants concerning his mistress: "The silly fool of a woman does not even know how to make children." The gibe being repeated to her, she exclaimed mockingly, "It's not my fault, you know." When Bonaparte heard of the rejoinder, his face clouded (10, p. 129). Although his first wife, Josephine, bore two children by a previous husband, she was unable to give Napoleon an heir; for this reason, Napoleon divorced her. For years his strongest desire was for an heir. Never-

theless, he is generally alleged to have been the father of three children.

Whether Napoleon sired three children or none, the fact remains that he was in sexual development either defective or near defective. What exact relationship this may have had to his genius we do not know. It obviously does not go to substantiate Terman's assertion that the eccentricity of genius is largely a myth. Certainly, the masculinity index of Napoleon would not have been high.

One more characteristic of Napoleon is worthy of mention. Napoleon had some kind of falling sickness which some have interpreted to be epilepsy. It is possible that his seizures would never have been recorded had not Talleyrand broken his vow to the emperor. Clark quotes Talleyrand as follows:

"Perhaps the most accurate account of an actual fit (epileptic) in Napoleon is given by Talleyrand,* who, in September, 1805, received instructions to accompany him to Strasburg. The day of Napoleon's departure from that city Talleyrand had dinner with the Emperor, and on leaving the table Napoleon went to see the Empress. He had only been with her a few minutes when suddenly he came out of her apartment, went to Talleyrand and took him by the arm and led him into his room. M. de Remussuat, the first chamberlain, entered at the same time and they had barely reached the Emperor's room when he fell to the floor. He barely had time to ask that the door be closed. Talleyrand tore away Napoleon's cravat, as it seemed to choke him. He groaned and foamed at the mouth, and had convulsions which ceased after a quarter of an hour. He was then placed in a chair and was able to speak. When he had dressed himself he enjoined secrecy on the two who had witnessed his seizure. Half an hour later he was on his way to Carlsruhe, and on reaching Stuttgart he wrote to Talleyrand, ending his letter with the words, 'I am well'" (3, p. 379).

An interesting footnote to the foregoing comment is added.

"As an aside it might be interesting to note here that the Battle of Austerlitz which Napoleon was fond of calling the 'Battle of the Three Emperors,' might also be christened the 'Battle of the Three Epileptics.' The Russian Emperor Paul was epileptic and died an epileptic dement, while the Archduke Charles was also known to suffer from epileptic fits. Indeed,

*Talleyrand, *From Consul to Emperor*. Classic Memories. London, p. 330.

the indecisive battle of Aspern-Essling preceding the fateful disastrous battle of Wagram six weeks later was probably rendered so by a series of fits from which the Archduke suffered the day after the Battle of Aspern-Essling. The boastful Austrians even have a monument to the doughty Archduke commemorating him as the 'victor over Napoleon'" (3, p. 381).

The question might be asked as to why Napoleon had epileptic seizures so infrequently. In reply to this query, Clark states that with certain epileptic subjects the attacks are infrequent. Moreover, Napoleon was unusually successful for many years in satisfying certain strong desires, and this success may have afforded an outlet for his energies, and thus have forestalled the attacks (3, pp. 382 f.) The attacks may have been much more frequent than one would suspect from the biographies, for Napoleon was extremely desirous that his affliction remain undetected.

MacLaurin expresses doubt that Napoleon had epilepsy in the ordinary sense of that term, though he admits that he may have had some form of this affliction.

"If Napoleon really had epilepsy it was assuredly not the '*grand mal*' which helps to fill our asylums. It is just possible that '*petit mal*' may have been in the picture. This is a curious condition which manifests itself by momentary loss of consciousness; the patient may become suddenly dreamy and purposeless, and may perform curious involuntary actions—even crimes—while *apparently* conscious. When he recovers he knows nothing about what he has been doing, and may even resume the interrupted action which had occupied him at the moment of the seizure. Some such explanation may account for Napoleon's fits of furious passion, that seem to have been followed by periods of lethargy and vomiting. It is a sort of pleasing paradox—and mankind loves paradox—to say that supremely great men suffer from epilepsy. It was said of Julius Caesar, of St. Paul, and of Mohammed. These men are said to have suffered from 'falling sickness,' whatever that may have been; there are plenty of conditions which make men fall to the ground, without being epileptic. Meniere's disease, for instance. It is ridiculous to suppose that Julius Caesar and Napoleon—by common consent the two greatest of the sons of men—should have been subject to a disease which deteriorates the intellect.

"It is possible that some such trouble as '*petit mal*' may have been at the bottom of the curious stories of a certain listless

torpor that appears to have overcome Napoleon at critical moments in his later battles" (12, pp. 223-224).

It seems that, whether Napoleon had the ordinary form of epilepsy or whether he had what MacLaurin terms "*petit mal*," he was afflicted with some form of nervous instability that caused him at times to behave in a most unusual manner. One naturally is led to inquire whether the congenital weakness exerted any influence upon his military activity. Regarding such an issue, one can only speculate. Clark (4) attributed Napoleon's military genius to the mental state attending epilepsy. Undoubtedly, in advancing this explanation, Clark has yielded to the tendency to over-simplify a very complex problem. Multitudinous elements were undoubtedly contributing factors in effecting the drive for which Napoleon remains an immortal. Neither a high IQ nor the presence of a physical handicap can be judged rightly as the cause of Napoleon's attainment. It is, however, entirely plausible that his falling sickness was *one* factor of great importance in determining the direction of his energy. Particularly likely is it that a capable man living *at a time when war was inevitable* would choose a military life as a possible means of escape from realization of his physical weakness. The physical weakness probably was a decisive factor in effecting Napoleon's choice of activity.

That Napoleon rose above the multitude was doubtless due in part to superior mental ability. His success was probably due also to a persistent and impelling desire to force the world to acknowledge his strength. The bitter and omnipresent knowledge of his physical afflictions undoubtedly had much to do with the establishment and maintenance of a desire to force recognition of his superiority. The old problem of causation again arises! But certainly neither a simple formula (IQ), nor a physical weakness, is sufficient to explain the attainment of an organism so complex as man. The fact seems apparent, however, that Napoleon was driven to attain world renown by deep-seated feelings of inferiority and by a knowledge of his physical weaknesses.

In the foregoing discussion the writers have pointed to several renowned military leaders and statesmen who were eccentric; some of them nervously unstable. It seems reasonable to infer that in certain instances the instability was a causal factor in effecting leadership. This hypothesis, of course, cannot be proved scientifically. Nevertheless, nervous instability probably produces certain subtle and

obscure drives. For the drive associated with the "falling sickness" of Napoleon and of Caesar, military endeavor might have been a resultant expression. Consequently, it is not surprising that other capable leaders have been afflicted in this manner. The ability is present, the nervous instability furnishes (in part at least) the drive, the result is military eminence.

Whether Caesar, Napoleon, Archduke Charles, or the Russian Emperor Paul, etc., would have attained eminence had they possessed the same potentialities and lacked the nervous instability is unknown. Surely, the hypothesis is justified that the instability of these men was *one* important causal factor. And, certainly, in any fairly complete analysis of causation, there should be an attempt to evaluate, sanely and impartially, the numerous subtle, elusive, but extremely important drives to human action. Too often are we satisfied when we have ascertained certain easily discernible causes of human action. Thus, we are sometimes prone to attribute too much importance to intelligence (as measured or estimated by present-day technique) when we attempt to evaluate the potential or proven genius.

REFERENCES

1. ARETZ, G. Napoleon and his women friends. Philadelphia: J. Lippincott, 1927. Pp. 375.
2. CARLYLE, T. On heroes, hero-worship, and the heroic in study. Boston: Houghton Mifflin, 1907. Pp. xxxvi+375.
3. CLARK, L. P. A Psycho-historical study of the epileptic personality in the genius. *Psychoanal. Rev.*, 1922, 9, 367-402.
4. ———. Napoleon, self-destroyed. New York: Cape & Smith, 1929. Pp. 253.
5. ———. Unconscious motives underlying the personalities of great statesmen and their relation to epoch-making events: I. A psychologic study of Abraham Lincoln. *Psychoanal. Rev.*, 1921, 8, 1-21.
6. COX, C. M., *et al.* Genetic studies of genius: Vol. II. The early mental traits of three hundred geniuses. Stanford University, Calif.: Stanford Univ. Press, 1926. Pp. xxiii+842.
7. DRYDEN, J., *et al.* Plutarch's lives of illustrious men. Vol. II. (Trans. from the Greek by John Dryden and others.) Boston: DeWolfe, Fiske & Co.
8. FENTON, J. C. Book review of Terman's *Genetic Studies of Genius*, Vol. I. *J. Appl. Psychol.*, 1927, 11, 174.
9. HERNDON, W. H., & WEIK, J. W. Abraham Lincoln: The true story of a great life. (2 vols.) New York: Appleton, 1893.
10. LUDWIG, E. Napoleon. New York: Boni & Liveright, 1926. Pp. xix+707.
11. MACLAURIN, C. Mere mortals: medico-historical essays. New York: Doran, 1925. Pp. xiii+291.

12. ————. *Post mortem: essays, historical and medical.* New York: Doran, 1922. Pp. vi+260.
13. SLOANE, W. M. *The Life of Napoleon Bonaparte.* (4 vols.) New York: Macmillan, 1915, 1. Pp. xiii+457.
14. SUTONIUS, T. C. *The lives of the twelve Caesars.* (Trans. by Alexander Thomson.) (Rev. and corrected by T. Forester.) London: George Bell, 1887. Pp. 557.
15. TARBELL, I. M. *A life of Napoleon Bonaparte.* New York: Macmillan, 1923. Pp. 485.
16. Terman, L. M. [Ed.] *The Twenty-third Yearbook of the National Society for the Study of Education.* Bloomington, Ill.: Pub. School Publ. Co., 1924, Pt. I. Pp. 444.
17. THADDEUS, V. *Julius Caesar and the grandeur that was Rome.* New York: Brentano's, 1927. Pp. 321.
18. VAN DOREN, C. *Swift.* New York: Viking Press, 1930. Pp. 279.
19. WITTY, P. A., & LEHMAN, H. C. Ability versus effective ability. *Psychol. Rev.*, 1928, **35**, 67-86.
20. ————. Drive: a neglected trait in the study of the gifted. *Psychol. Rev.*, 1927, **34**, 364-376.
21. ————. Nervous instability and genius: poetry and fiction. *J. Abn. & Soc. Psychol.*, 1929, **24**, 77-90.
22. ————. Religious leadership and stability. *Psychol. Rev.*, 1929, **36**, 56-82.

Northwestern University
Evanston, Illinois

Ohio University
Athens, Ohio

L'INSTABILITÉ NERVEUSE ET LE GÉNIE: GRANDS HOMMES MILITAIRES ET POLITIQUES

(Résumé)

Dans cet article il s'agit de plusieurs exemples du génie militaire et politique chez lesquels le manque de la capacité de se soumettre aux contraintes dans la jeunesse aurait eu comme résultat une impulsion persistante peu commune. Cette impulsion, combinée avec une quantité raisonnable d'habileté et d'opportunité, a fait atteindre le rang de génie reconnu à ces individus. L'habileté existe, l'instabilité nerveuse donne (en partie, au moins) l'impulsion, le résultat est l'éminence militaire. Les auteurs ne disent pas que les génies militaire et politique sont accompagnés invariablement d'un dérangement nerveux. Cependant, ils croient, après l'étude de grands hommes tels que Cromwell, Frédéric le Grand, Lincoln, Césaire, et Napoléon, que les grands accomplissements militaires et politiques ont quelquefois été des moyens très satisfaisants d'échapper à la torture mentale. On ne sait si ces hommes auraient atteint la même éminence s'ils avaient possédé les mêmes capacités potentielles et manqué de l'instabilité nerveuse. On peut certainement justifier l'hypothèse que l'instabilité de ces hommes a été un facteur causal important. Certainement, dans n'importe quelle assez complète analyse de la causation, on devrait essayer d'évaluer, d'une façon saine et impartiale, les nombreuses impulsions à l'action humaine, subtiles,

trompeuses, mais un peu plus importantes. On tend quelquefois à attribuer trop d'importance à l'intelligence (comme mesurée ou estimée par la technique de nos jours) quand on essaie d'évaluer le génie potentiel ou prouvé.

WITTY ET LEHMAN

NERVÖSE LABILITÄT (INSTABILITY) UND GENIE: MILITÄRISCHE UND POLITISCHE FÜHRER

(Referat)

Die Verfasser behandeln in diesem Aufsatz mehrere Beispiele des Militärischen und politischen Genies, worin die Unfähigkeit, Durchkreuzung (thwarting) in der Kindheit auszulasten, sich als ungewöhnlicher und aushaltender Drang ausgewirkt zu haben scheint. Dieser Drang, mit mässiger Tüchtigkeit (ability) und Gelegenheit verbunden, hat es diesen Individuen möglich gemacht, den Rang des anerkannten Genies zu erreichen. Es besteht die Tüchtigkeit, die nervöse Labilität bereitet wenigstens teilweise den Drang, und es ergibt sich das militärische Hervorragende (military eminence). Die Verfasser gehen nicht so weit, dass sie behaupten würden, dass das Genie immer von einer nervösen Störung begleitet sei. Sie glauben aber doch, nach genauer Untersuchung von Führern wie Cromwell, Friedrich der Grosse, Lincoln, Cäsar und Napoleon, das hervorragende militärische und politische Leistungen manchmal als höchst befriedigende Bahnen zur Entfaltung des geistigen Druckes (stress) gewirkt haben. Ob die obengenannten Personen die gleiche hohe Stelle erreicht hätten, wenn ihnen die selben Fähigkeiten eigen gewesen wären und die nervöse Labilität nicht bestanden hätte, bleibt unbekannt. Es ist aber sicher die Hypothese berechtigt, dass die Labilität *eine* wichtige wirkende Ursache gewesen ist. Und gewiss soll bei einer irgendwie vollständigen Analyse der Verursachung der Versuch gemacht werden, die zahlreichen feinen, ausweichenden aber sehr wichtigen Dränge auf vernünftige und unparteiische Weise zu bewerten. Wenn wir versuchen, dass potenzielle oder erwiesene Genie zur bewerten, so sind wir manchmal geneigt, der Intelligenz (so wie sie mit den gegenwärtigen Methoden gemessen oder abgeschätzt wird) eine zu grosse Wichtigkeit beizumessen.

WITTY UND LEHMAN

SHORT ARTICLES AND NOTES

THE INFLUENCE OF LANGUAGE BACKGROUND ON INTELLIGENCE TESTS

RUDOLF PINTNER

One underlying assumption of all intelligence testing is that the background of the individuals to be compared is roughly similar with respect to the situations, tasks, or tests presented to them, by means of which we measure the differences in intelligence among such individuals. It is a debatable question as to how accurate our intelligence measurements are with reference to children in this country living in homes where a foreign language is spoken. Some writers believe this bilingual environment has a negligible influence on the results of the usual verbal group or individual intelligence tests, while others believe it has considerable influence upon such tests. The amount of such influence may very well vary from a great deal to very little, depending upon the amount of foreign language used in the child's home. It may also vary with the age of the child. The younger the child the more dependent is he likely to be upon the narrower environment of the home. Correlations of verbal intelligence test scores with school achievement can tell us nothing about the validity of the intelligence test scores for such children, because any language handicap that is influencing the intelligence test scores will also be present in the child's school achievement.

The present article deals with only one small part of the problem of language handicap or of the larger problem of bilingualism in general. A comparison has been made of the scores on two tests of children from English-speaking homes and of children from homes in which another language may have been used in addition to English. These two groups will be called the English and non-English, remembering, of course, that the non-English were all living in the English-language environment of an American city and came from homes in which all different amounts of English may have been used. First-grade children in three schools in New York City were tested.¹ School A was a large school in the Bronx, and the non-English in this school consisted largely of Italian children. The division between English and non-English was made solely on the basis of the

¹The writer wishes to thank the three principals, Miss Ellen M. Phillips, Mr. Burt P. Seelye, and Mr. Abraham Cohen, for their splendid assistance and cooperation in this work. He also wishes to thank his students for their help in this study and especially Miss Edith Potts, Mr. George Forlano, Mr. Alexander Taylor, Mr. Jacob Rabinowitz, and Dr. Karel Prüg.

child's name. All Anglo-Saxon looking names were put into one group and all foreign names into the other. School B was a school on the upper East side of Manhattan. Here the foreign group consisted of a fair number of Bohemian children and these were carefully selected by Dr. Prügl, a Bohemian himself. The others were sorted out according to name. School C was a school on the lower East side of Manhattan. Here the teachers were asked to divide the children. They divided the 164 children into two groups; one group of 71 children coming from homes in which no English was spoken and another of 93 in which English was spoken. A study of the children's names, however, revealed only a dozen or so Anglo-Saxon names, the rest being predominantly Jewish. In all probability, therefore, practically all the children in this school came from bilingual homes. No real differentiation between English and non-English was possible. We shall, however, present the results for the two divisions as made by the teachers.

The tests used were the Pintner-Cunningham Primary Mental Test (1, 2) and the Pintner Primary Non-Language Test (3). The former is non-verbal in content and verbal in directions. The latter is entirely non-verbal throughout, the directions being given by means of pantomime and examples on the blackboard. A child from a foreign-language-speaking home could in no way be handicapped on the second test. On the first test he would be handicapped to the extent that he could not follow the verbal directions given by the examiner, but no knowledge of writing or reading is presupposed by the test. Both tests have been found to be fairly reliable and valid. They correlate from .5 to .7 with each other in various grade groups.

All the children in Grades IA and IB were tested in all three schools. Only children who were present during both of the tests were included in this study. The numbers are as follows:

School	English	Non-English	Total
A	57	112	169
B	30	67	97
C	93	71	164

If the bilingual child is handicapped on language tests, then he should do relatively better on non-language tests. When his position on a language test is compared with his position on a non-language test, he should show up better on the non-language test. Our best method of testing this hypothesis will be to compare the differences between the mean scores of the two groups of children and express these differences in terms of the standard deviation of either of the groups. Furthermore, we shall keep the three schools separate, so that the English and non-English living in the same section of the city will be compared. The essential data are given in Table 1.

TABLE 1
MEAN DIFFERENCES BETWEEN ENGLISH AND NON-ENGLISH GROUPS

School	<i>English</i>		<i>Non-English</i>		Diff. of Means	<i>Diff.</i>	<i>Diff.</i>
	Mean	Sigma	Mean	Sigma		σ Eng.	σ Non-Eng.
<i>Pintner-Cunningham Test</i>							
A	24.96	9.31	16.63	7.46	—8.33	—0.895	—1.12
B	27.50	7.04	29.10	8.76	+1.6	+0.227	+0.183
C	23.51	9.90	19.80	8.40	—3.71	—0.374	—0.441
<i>Pintner Primary Non-Language Test</i>							
A	44.37	12.79	40.35	11.12	—4.02	—0.314	—0.359
B	44.50	10.84	49.54	10.50	+5.04	+0.465	+0.480
C	49.71	12.66	44.62	12.18	—5.09	—0.402	—0.418

In School A we find that the non-English group scores 8.33 points less than the English on the Pintner-Cunningham Test. In terms of the *S.D.* of the English group, this means they are .89 σ below the English group. Turning now to the same groups on the Non-Language Test, we find that the non-English group scores 4.02 points less than the English group. Again, in terms of the *S.D.* of the English group, this means they are .31 σ below the English group. The non-English group has accomplished relatively more on the Non-Language Test. Comparing the groups in terms of the *S.D.* of the non-English group, we find the same tendency of the non-English group to score relatively higher on the Non-Language Test.

In School B we find that the non-English group exceeds the English group on both tests. On the Pintner-Cunningham the non-English are .23 σ above the English, and on the Non-Language Test they are .46 σ above the English. They show a greater superiority to the English on the Non-Language Test. In both schools, therefore, the Non-Language test allows the non-English group to score higher than does the Pintner-Cunningham. In School A this is shown by the non-English group being less inferior on the Non-Language Test, and in School B by being more superior on this test. Or, in other words, the Pintner-Cunningham Test increases the inferiority of the non-English group in School A and decreases their superiority in School B.

When we turn to School C, made up almost entirely of foreign children divided into two groups by the teachers, we find the non-English group to be inferior to the English in both tests, and in both cases to about the same amount. The non-English are below the English .37 σ on the Pintner-Cunningham and by .40 σ on the Non-Language Test. The so-called non-English group in this school has no advantage on the Non-Language Test. In reality we have no real division into English and non-English in this school. What we seem to have is a very rough division into slightly inferior and slightly superior groups. The teachers may have inferred the amount of language handicap from the general achievement of the child in school.

Another method of comparing the English and non-English groups is to find the percentage of the one group reaching or exceeding the mean of the other group. Below we give the percentage of the non-English group reaching or exceeding the mean of the English group on the two tests:

School	Percentages	
	P-C Test	N-L Test
A	13	36
B	59	67
C	31	37

In all schools the percentages on the Non-Language Test are larger than on the Pintner-Cunningham Test. The greatest difference here is found in School A and the least in School C. All the evidence again is in favor of the greater relative superiority of the non-English on the Non-Language Test.

If we enquire as to whether the differences of the mean scores of the various groups are statistically significant, we have the results shown in

TABLE 2
SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE MEANS

School	<i>Mean</i>		<i>Sigmas of Means</i>		Sigma of Diff. of means	<i>Diff.</i> σ Diff.
	English	Non-English	English	Non-English		
<i>Pintner-Cunningham Test</i>						
A	24.96	16.63	1.23	0.71	1.418	5.87
B	27.5	29.1	1.28	1.06	1.64	0.97
C	23.51	19.80	1.026	0.99	1.425	2.60
<i>Non-Language Primary Test</i>						
A	44.37	40.35	1.71	1.046	1.337	3.01
B	44.5	49.5	1.98	1.26	1.83	2.73
C	49.71	44.62	1.31	1.44	1.945	2.62

Table 2. We note significant differences in both tests in School A, with a smaller ratio for the Non-Language Test, where the non-English group comes up closer to the English. In School B the difference between the two groups is not significant on the Pintner-Cunningham Test, but it is much more so on the Non-Language Test, where the non-English group is able to show its superiority. In School C the ratios are the same for both tests, just as we should expect.

We present, finally, the correlations between the two tests for the various groups:

School	English	Non-English	All children
A	.55	.25	.42
B	.45	.56	.53
C	.54	.55	.58

We find here no general tendency for the correlations to be higher in the English rather than in the non-English groups.

In School B we found the non-English group superior to the English not only on the Non-Language Test, but also on the Pintner-Cunningham Test. In this school the non-English group was further divided into 28 Bohemians and 39 other foreign children. The mean scores on the Pintner-Cunningham for the English, Bohemian, and foreign are 27.5, 30.3, and 28.5, respectively. The English group is lowest and the Bohemian highest. The superiority of the Bohemian is .4 σ above the English mean; the superiority of the other foreign is only about .1 σ above the English mean. On the Non-Language Test the mean scores for the groups are: English 44.5; Bohemian 48.5; Foreign 49.5. Here the superiority of the Bohemian is .37 σ above the English mean and that of the foreign is .46 σ above the English mean. The foreign or non-Bohemian group profits most from the Non-Language Test. The Bohemian group is as superior to the English on the Pintner-Cunningham as it is on the Non-Language. Evidently the amount of language required to understand the directions of the Pintner-Cunningham Test is no particular handicap to this Bohemian group.

Summary. Four hundred and thirty children in Grade 1 in three New York schools were divided into groups having English- or Non-English-speaking home backgrounds. They were given the Pintner-Cunningham and the Pintner Non-Language Primary Tests. In two of the schools the non-English groups did relatively better on the Non-Language Test. They seemed to be handicapped on the Pintner-Cunningham Test, and this in spite of the fact that the non-English-speaking group in one school exceeded the English on this test. In the third school no differences between the groups on the two tests were found, and this is probably due to the fact that no real division between English and non-English was made.

It would seem from this study that great caution should be exercised in the comparison of children with different language backgrounds when they are being compared by means of verbal intelligence tests. It should be remembered that the Pintner-Cunningham Test is non-verbal in content and that only enough English to understand the directions is required. With tests verbal in content as well as directions a greater handicap for non-English children is probable. On the other hand, the language handicap will presumably be greatest in Grade 1, diminishing thereafter as the child grows older. At what grade this language handicap will be entirely overcome

will depend upon many factors, notably the opportunity to mix in an English-speaking environment and the general intelligence of the individual. A bilingual environment may prevent some individuals from ever really indicating their maximum intelligence on a verbal group intelligence test.

REFERENCES

1. PINTNER, R., & CUNNINGHAM, B. V. The problem of group intelligence tests for very young children. *J. Educ. Psychol.*, 1922, **13**, 465-472.
2. ———. The Pintner-Cunningham Primary Test. *J. Educ. Psychol.*, 1927, **18**, 52-58.
3. ———. Non-Language Primary Mental Test. New York: Bur. Publ., Teach. Coll., Columbia Univ., 1930.

*Teachers College
Columbia University
New York City*

CONCERNING THE THURSTONE "PERSONALITY SCHEDULE"

O. L. HARVEY

The Thurstone Personality Schedule, 1929 edition, is based upon material reported in the first number of the *Journal of Social Psychology*. According to that article, the procedure of schedule construction was as follows:

- (a) From textbook summaries of psychoneurotic symptoms were compiled some 600 and more statements having reference to psychoneurotic behavior;
- (b) These statements were cast into subjectively determined categories (unfortunately not described);
- (c) Duplicates were eliminated in so far as possible, with the exception, however, of items which constituted duplicates only in the pattern of behavior involved, but not in the actual situation in which it is manifested (e.g. items 54, 163¹);
- (d) Every question was then checked either "yes" or "no" in the way in which it was expected that psychoneurotic subjects would probably answer it;
- (e) The list of items remaining (number of items not indicated by author) was administered to some 694 freshmen at the University of Chicago;
- (f) It was found that the distribution of total scores yielded a positively skewed biometric distribution.
- (g) Validity could be established against no satisfactory external criterion, so the criterion of internal consistency was adopted. The 50 subjects (sex not indicated) receiving the

¹These numbers refer to the order in the 1929 published schedule.

topmost total scores (i.e., the most neurotic subjects, according to the test) were compared with the bottommost group of 50 least neurotic subjects. The proportions of each group responding neurotically were compared for every item separately. In every case, but one, it was found that the statistical findings confirmed the guess of the schedule-constructors. In every such case the proportion of "neurotic" subjects responding neurotically (in the way described in the textbooks) was greater than the proportion of "well-adjusted" subjects responding neurotically. These facts were adduced as adequate, or at least as the best available evidence for thinking the schedule sufficiently discriminative as an instrument.

(h) A special list of the 42 items yielding the greatest degree of discrimination was printed in full.

Interested in the schedule, but not entirely satisfied with it, the writer administered it to 146 male students taking a required course in business psychology at the University of Texas. The distribution of total scores obtained by this group is compared in Table 1 with those reported for men by Thurstone. It is evident that the Texas group is somewhat more neurotic and more variable than the Chicago group. This might be due to the fact that the Texas group represents predominantly seniors and

TABLE 1

DISTRIBUTION OF TOTAL SCORES ON THE THURSTONE PERSONALITY SCHEDULE
COMPARING EXPERIMENTAL GROUP (TEXAS) WITH STANDARDIZATION
GROUP (CHICAGO)

Scores	Frequency totals		Frequency percentages				Letter grades
	Chicago	Texas	Chicago	Texas			
0- 9	35	8	9.2	5.5	} 50.5	} 33.6	A and B
10- 19	79	19	20.8	13.0			
20- 29	78	22	20.5	15.1			
30- 39	73	27	19.2	18.5	} 39.0	} 41.1	C
40- 49	47	22	12.4	15.1			
50- 59	27	11	7.2	7.5			
60- 69	15	15	3.9	10.3	} 10.9	} 25.4	D and E
70- 79	12	10	3.2	6.8			
80- 89	9	5	2.4	3.4			
90- 99	4	2	1.1	1.4			
100-109	1	2	0.3	1.4			
110-119	0	1	0.0	0.7			
120-129	0	2	0.0	1.4			
	380	146	100.2	100.1			

TABLE 2
THE TEXAS DATA

Showing (a) frequency of psychoneurotic responses per item for the total Texas group of 146 males (*N*); (b) percentage of psychoneurotic responses per item on each subgroup of 49 AB, 60 C, and 37 DE letter-grade subjects; and (c) differences between percentages of psychoneurotic responses for the DE subgroup compared with the AB subgroup.

Item	<i>N</i>	AB(%)	C(%)	DE(%)	DE(%) - AB(%)
1	33	8	23	40	32
2	15	4	7	24	20
3	88	53	60	70	17
4	19	8	12	22	14
5	17	2	15	19	17
6	11	0	3	24	24
7	9	4	7	8	4
8	11	4	3	24	20
9	15	4	12	16	12
10	59	31	40	54	23
11	43	12	27	57	45
12	37	12	27	40	18
13	43	14	32	46	32
14	11	2	7	16	14
15	20	10	15	16	6
16	74	12	60	87	75
17	25	10	18	24	14
18	73	37	57	57	20
19	64	27	50	57	30
20	44	16	33	43	27
21	67	12	50	84	72
22	11	4	7	14	10
23	42	10	27	57	47
24	50	16	33	60	44
25	21	4	15	27	23
26	4	2	2	5	3
27	3	0	2	5	5
28	56	12	38	73	61
29	3	2	2	3	1
30	39	6	33	43	37
31	62	14	58	54	40
32	55	18	37	65	47
33	28	12	15	35	23
34	22	4	17	27	23
35	35	2	28	46	44
36	11	2	7	16	14
37	16	2	13	19	17
38	17	4	12	22	18
39	34	4	22	52	48
40	19	2	17	22	20
41	27	12	17	30	18
42	44	18	28	49	31
43	27	8	18	32	24
44	22	6	10	35	29
45	14	2	10	19	17

TABLE 2 (*continued*)

Item	N	AB(%)	C(%)	DE(%)	DE(%) - AB(%)
46	16	0	10	27	27
47	35	6	22	52	46
48	40	8	27	54	46
49	56	23	38	60	37
50	35	12	27	35	23
51	4	0	0	11	11
52	67	21	45	81	60
53	32	12	20	38	26
54	47	12	35	54	42
55	29	10	22	30	20
56	1	0	2	0	0
57	48	12	37	54	42
58	20	2	13	30	28
59	14	2	5	27	25
60	0	0	0	0	0
61	14	0	5	30	28
62	26	6	17	35	29
63	0	0	0	0	0
64	10	2	8	11	9
65	10	4	8	8	4
66	25	2	22	30	28
67	11	0	8	16	16
68	22	2	10	40	38
69	12	4	5	19	15
70	50	21	40	43	22
71	12	2	7	19	17
72	53	23	35	57	34
73	86	35	65	81	46
74	12	0	8	19	19
75	47	23	38	35	12
76	19	4	13	24	20
77	33	16	23	30	14
78	56	23	38	60	37
79	45	16	27	57	41
80	78	33	52	84	51
81	31	18	18	30	12
82	9	2	5	14	12
83	22	0	18	27	27
84	6	0	2	14	14
85	66	27	45	70	43
86	40	10	33	40	30
87	4	0	3	5	5
88	22	8	17	22	14
89	61	21	42	70	49
90	27	10	17	32	22
91	53	25	37	52	27
92	36	12	32	30	18
93	10	2	5	16	14
94	54	16	38	62	46
95	38	14	27	40	26
96	31	2	25	40	38

TABLE 2 (*continued*)

Item	N	AB(%)	C(%)	DE(%)	DE(%) - AB(%)
97	9	2	5	14	12
98	43	8	32	54	46
99	31	2	17	54	52
100	26	2	12	49	47
101	3	0	2	5	5
102	2	2	2	0	-2
103	6	0	7	5	5
104	9	0	3	19	19
105	14	2	8	22	20
106	9	2	3	16	14
107	5	0	5	5	5
108	40	4	28	57	53
109	14	4	8	19	15
110	49	12	35	60	48
111	25	8	15	32	24
112	16	2	5	32	30
113	35	10	25	40	30
114	12	6	7	14	8
115	37	6	25	52	46
116	3	0	2	5	5
117	29	0	20	46	46
118	15	0	12	22	22
119	9	0	5	16	16
120	34	2	23	54	52
121	29	12	22	27	15
122	8	4	10	0	-4
123	33	2	22	52	50
124	80	39	58	70	31
125	11	4	7	14	10
126	21	2	12	35	33
127	41	6	25	62	56
128	34	10	23	40	30
129	10	2	2	22	20
130	34	10	27	35	25
131	26	8	20	27	19
132	3	0	3	3	3
133	17	0	10	30	30
134	4	0	7	0	0
135	6	0	3	11	11
136	29	4	23	35	31
137	32	10	23	35	25
138	59	14	45	68	54
139	53	8	37	73	65
140	30	4	17	49	45
141	7	2	3	11	9
142	10	0	2	24	24
143	26	6	15	46	40
144	4	0	3	5	5
145	67	29	50	62	33
146	51	25	38	46	21
147	16	6	7	24	18

TABLE 2 (*continued*)

Item	N	AB(%)	C(%)	DE(%)	DE(%) - AB(%)
148	47	14	25	68	54
149	26	6	20	30	24
150	69	31	53	60	29
151	51	8	32	76	68
152	45	4	23	79	75
153	39	4	25	60	56
154	21	6	8	35	29
155	12	0	12	14	14
156	22	6	20	19	13
157	19	0	15	27	27
158	22	0	12	40	40
159	46	18	30	52	34
160	7	0	5	11	11
161	72	35	50	68	33
162	19	2	13	27	25
163	36	4	22	57	53
164	37	14	18	52	38
165	3	2	2	3	1
166	5	0	3	8	8
167	46	23	38	32	9
168	42	6	23	68	62
169	49	18	35	52	34
170	7	4	2	11	7
171	1	0	0	3	3
172	31	4	23	40	36
173	16	4	8	24	20
174	31	4	15	54	50
175	46	18	37	40	22
176	15	4	12	16	12
177	66	12	47	87	75
178	31	2	17	54	52
179	44	4	35	57	53
180	45	14	33	49	35
181	73	27	53	76	49
182	34	2	22	54	52
183	40	16	30	38	22
184	50	12	38	57	45
185	6	2	2	11	9
186	31	4	17	52	48
187	70	37	48	62	25
188	35	12	28	32	20
189	14	0	13	16	16
190	73	47	55	46	-1
191	17	0	10	30	30
192	53	14	30	76	62
193	37	6	20	60	54
194	10	0	10	11	11
195	66	18	47	79	61
196	20	4	17	22	18
197	27	8	15	38	30
198	1	0	2	0	0

TABLE 2 (*continued*)

Item	N	AB(%)	C(%)	DE(%)	DE(%) - AB(%)
199	8	0	3	16	16
200	18	8	10	22	14
201	45	4	30	68	64
202	51	6	33	76	70
203	4	0	0	11	11
204	21	4	15	27	23
205	42	8	37	43	35
206	59	14	50	60	46
207	24	8	13	32	24
208	29	8	15	43	35
209	21	8	15	22	14
210	6	0	3	11	11
211	2	0	0	5	5
212	13	4	7	24	20
213	5	0	3	11	11
214	40	21	40	62	40
215	6	4	3	8	4
216	1	0	0	3	3
217	1	0	0	3	3
218	8	2	20	11	9
219	1	0	0	3	3
220	29	14	18	43	29
221	17	0	18	32	32
222	22	2	32	49	47
223	9	8	8	14	6

sophomores, from a school of business administration, in a Southern University. However, differences which later tables will sustain are to be considered in the light of this evidence.

The number of psychoneurotic responses to every item (using the scoring device supplied with the schedule—a device different from that used in Thurstone's original study in that now only "yes" and "no" answers are scored) is reported in Table 2. (Unfortunately, Thurstone provided no such complete table for the Chicago group.) The A and B letter-grade groups (Thurstone's classification) were combined in a group of 49 subjects, as against the D and E combined group of 37 subjects. For each of the three groups thus formed (i.e., AB, C, DE), and for every item, was calculated the percentage of psychoneurotic responses recorded, together with the difference between the DE and the AB percentages. The two most important columns in this table thus become (*a*) the column (*N*) of actual frequencies of neurotic responses for the total group of 146 subjects, and (*b*) the column of differences between the AB and DE groups. The first provides a measure of the degree to which the item itself prevails as a psychoneurotic symptom in a relatively unselected population. Obviously,

TABLE 3

SOME COMPARISONS OF THE THURSTONE AND HARVEY DATA

Comparison of (a) topmost 42 discriminative items according to Thurstone, with (b) topmost 41 frequency items according to Harvey, and with (c) topmost 45 discriminative items according to Harvey, showing to what extent, and with respect to which items, these three lists of items are in agreement.

Item	Group (a)	Group (b)	Group (c)	Item	Group (a)	Group (b)	Group (c)	Item	Group (a)	Group (b)	Group (c)
3	x	x		80		x	x	152	x		x
10		x		85	x	x		153	x		x
13	x			89		x	x	161	x	x	
15	x			90	x			163			x
16		x	x	91		x		168	x		x
18		x		94	x	x	x	169		x	
19		x		98	x		x	174			x
21	x	x	x	99	x		x	177	x	x	x
23	x		x	100	x		x	178			x
24		x		108	x		x	179			x
28	x	x	x	110		x	x	181	x	x	x
30	x			115			x	182			x
31	x	x		117			x	184	x	x	
32		x	x	120			x	186			x
39	x		x	123			x	187	x	x	
47	x		x	124	x	x		190		x	
48			x	127	x		x	192		x	x
49		x		136	x			193			x
52	x	x	x	138		x	x	195	x	x	x
57	x	x		139	x	x	x	196	x		
70		x		145	x	x		197	x		
72		x		146		x		201	x		x
73	x	x	x	148			x	202	x	x	x
78	x	x		150		x		206	x	x	x
79	x			151		x	x	220	x		
								222			x

if low, it cannot be a very characteristic neurotic symptom. The second provides, as it were, a diagnostic measure—a measure of differentiation between the mentally better and the mentally worse adjusted. Other things being equal, the greater this difference, the more significant the symptom. Thurstone used only the criterion of percentage differences. It would seem, however, that not only one but both of these measures are relevant and therefore essential in determining the significance of an item. Unfortunately, Thurstone's omission makes an investigation of this aspect of the problem impossible. In any case, the writer found, with respect to his own group that high differentiation and high frequency do not necessarily go hand in hand. The 57 items having highest frequencies of total neurotic

response (score: 43 plus) were checked against the 58 items yielding the highest discriminatory differences (40 plus); 32 were found to be common to both modalities. It is these 32 items, thus selected, which constitute, in the opinion of the writer, the most significant in the schedule.²

Let us now turn to a closer examination of our selected 32 most significant items. Items 206 and 21 are eliminated from the list on the ground that they duplicate very closely items 52 and 177, respectively, the latter already being included in the list, and having better statistical claim to retention. Item 31 is eliminated because of its peculiarly distorted percentage progression in passing through C group from the AB to the DE group. Item 161, having the next highest frequency and discrimination, serves to round out the list, thus depleted, to make up a total of 30 items, as follows:

11, 16, 24, 28, 32, 52, 54, 57, 73, 79, 80, 85, 89, 94, 98, 110, 138,
139, 148, 151, 152, 161, 177, 179, 181, 184, 192, 195, 201, 202.

An examination of these items suggests the same conclusion as that which Thurstone reached after examining his own abbreviated list, namely, that the psychoneurotic is "one whose imagination somehow fails to express itself effectively on social reality." It should be noted that in neither of these abbreviated scales do there appear any of what may be termed accidental symptoms—facts concerning which the subject is not directly responsible (e.g., items 87, 102, 188)—an observation which gives rise to the next criticism of the Thurstone schedule.

A close examination of the schedule leads to the suspicion that certain patterns of psychoneurotic behavior are unduly weighted as a result of the insertion of a relatively larger number of items characteristic of those patterns. Thurstone presumably saw this possibility, but did not treat with it in his article in any detail. The writer, however, was interested in discovering whether or not there might be something of importance in the weighting of items. He therefore cast the 223 items of the schedule into a few fairly broad, but fairly disparate, categories. That this classification

²It is interesting to note at this stage (see Table 3) that, comparing Thurstone's 42-topmost discriminatory items with the topmost 45 obtained by the writer, we find 23 items in common—more than a 50% agreement. Comparing Thurstone's list with the Texas list of 41 items of topmost *frequency*, we find 21 items in common—again about 50% agreement. But, comparing the writer's topmost 45 discriminatory items with his own topmost 41 frequencies, we find as few as 19 items in common. And, pooling Thurstone's topmost 42 in the one modality with the writer's topmost 19 in both modalities, there are only 11 items (of which two are mere duplicates) in common. That these differences are attributable to some extent to differences in sampling is probably fairly certain. It would, nevertheless, be interesting to see what agreement there would have been had Thurstone selected his group of most significant items on the basis of frequency as well as of discrimination.

TABLE 4

ALLOCATION OF ITEMS ACCORDING TO GENERAL CATEGORY

(Items which are printed in boldface type are duplicated in some other category.)

The connotations of the terms here used are as follows:

1. *Sex*—Items having reference to sex behavior or attitudes, and special reactions to persons of the same or opposite sex.

2. *Family*—Items having reference to one's own family history; one's relationships with and reactions concerning other members of one's family; the behavior of other members of the family.

3. *Childhood*—Items having any reference to childhood.

4. *Sleep*—Items having any reference to sleep.

5. *Physique*—Items having reference (other than metaphorical) to organs of the body, to one's own health, and to one's susceptibility to drugs, etc. Questions such as a physician might ask of his patient.

6. *Consistency*—Items having reference to variations in mood, to impulsiveness, instability, equanimity.

7. *Ascendancy*—(This term is probably not coterminous in connotation with that used by Allport. It seems to be so appropriate, however, that it is used here, with apologies.) Items having reference to one's *conduct* with respect to other persons, or to situations in which persons form the background; one's superiority or inferiority relative to others; self-consciousness (perhaps implying an awareness of social approval or disapproval, with the concomitant experience of self-elation or self-abasement); leadership.

8. *Drive*—Items referring to one's activity, *regardless of* social reactions; animality, planfulness, self-confidence; listlessness, weariness, boredom.

9. *Sociality*—Items having reference to one's *attitudinal* relationship to other persons; tendencies to withdraw from, to approach, to seek, to attack, to become entirely oblivious to, other persons.

0. *Impulsions*—Items having reference to fears, phobias, impulsions, obsessions, worries, apprehensions, nervousness.

1. *Sex*: 4, 22, 29, 68, 81, 95, 125, **128**, **130**, 137, 159, 171, 172, 222.

2. *Family*: 4, 8, 22, 45, 69, 77, 81, 86, 102, 107, 122, 134, 147, 159, 165, 188, 212, **217**, **221**.

3. *Childhood*: 1, 74, 77, 105, 157, 215, **218**.

4. *Sleep*: 50, **160**, **166**, 223.

5. *Physique*: 5, 7, 9, 27, 34, 37, 42, **48**, 56, 60, 65, 66, 78, 82, 87, 88, 93, 110, 132, **133**, 141, 144, 155, 176, 189, 191, 198, 204, **211**, 213, 216, **218**.

6. *Consistency*: 32, 39, 41, 52, 80, 85, 90, 124, 127, 136, 139, 142, 146, 152, 158, 178, 193, 206, 207, **221**.

7. *Ascendancy*: 3, 10, 12, 14, 18, 30, 57, **68**, 70, 91, 94, 96, 98, 121, 145, **150**, 161, 173, 175, 180, 181, **187**, 195, **196**, 201.

8. *Drive*: 2, 6, 15, 28, 38, 44, 46, **48**, 54, 59, 71, 79, **104**, 117, **133**, 135, 148, 162, 163, **166**, 167, 168, 174, 179, 182, 183, 184, 192, **196**, 202, **203**, 210, 214, 220.

9. *Sociality*: 1, 11, 13, 17, 19, 21, 25, 33, 35, 36, 40, 43, 61, 67, 72, 73, 83, 89, **97**, 99, 103, 108, 112, 113, 115, 116, 120, 123, 126, **128**, 131, 138, 140, 143, **150**, 151, **157**, 164, 169, **171**, **172**, 177, **187**, 205, 209, **211**, **217**, **222**.

0. *Impulsions*: 16, 23, 24, 26, 47, 49, 51, 53, 55, 58, 62, 63, 64, 75, 76, 84, 92, **97**, 99, 100, 101, **104**, 106, 109, 111, 114, 119, 129, **130**, 154, 156, **160**, 185, 186, 194, 197, 199, **203**, 219.

x. *Indeterminate*: 20, 31, 118, 149, 153, 170, 190, 200, 208.

was admittedly arbitrary, and is probably not very reliable, does not entirely invalidate the procedure; it happens that the majority of items actually are separable and classifiable in a rough way. There is some overlapping, particularly between the categories of "ascendance" and "drive"; and certain items (e.g., 68) had to be classified under two separate heads. On the whole, however, the writer found, after some experimentation, that the task of classification was not entirely impossible. The items were distributed into ten categories (plus one for indeterminate items) as indicated in Table 4.

In terms of the number of items per category, the rank order of categories in Table 4 is as follows:

Category:	9	0	8	5	7	6	2	1	3	4
Number of items:	48	39	34	32	25	20	19	14	7	4

Attention is simply drawn to the fact that "physique" gains a place above the median, and "consistency" below.

The 57 items having the topmost frequencies of total responses (score: 43 plus) were checked; the same was done to the 58 items yielding the highest discrimination (40% plus) between the DE and the AB groups.³ It is interesting, by comparison with the findings of the preceding paragraph, to note that the distribution of the sums of these check-marks yields a rank order of categories as follows:

Category:	9	7	8	6	0	5	1	2	4	3
Total check-marks:	21	17	16	13	10	4	2	1	0	0

Here one notes that the relatively unimportant categories are those relating to "childhood," "sleep," "family," "sex," and "physique." "Physique" and "consistency" have changed sides—as perhaps one would expect.

Finally, the abbreviated scale, discussed above, was developed. The rank order of categories, in terms of the distribution of these finally selected items, is as follows:

Category:	8	7	9	6	0	5	4	3	2	1
No. of selected items:	8	7	6	6	2	1	0	0	0	0

Note how "consistency," "ascendance," "drive," and "sociality" hold their ground; whereas "sex," "childhood," "family," "sleep," and "physique" drop out almost entirely. It is interesting to note, furthermore, that "impulsions" drops quite low.

In general, if these few last tables mean anything at all, it would seem that, granting the excellence of the Thurstone scale, it could, nevertheless, possibly be improved by giving greater weight to those categories which are more essentially related to the subject's own attitudes and emotional be-

³The 32 of these items which checked both ways were selected as the basis for the abbreviated scale of 30 items discussed above.

havior—e.g., by increasing the number of representative items, or by increasing the weight of certain individual items; and giving less weight to categories having reference to those incidental factors which are, relatively speaking, outside of the subject's control. That these latter items actually are to be found in the neurotic syndrome the writer does not attempt to dispute; but it is argued that, in any attempt to develop not merely a checklist of symptoms, but also a measure of the degree of neurosis, some consideration should be given to the relative weighting of the symptoms manifest. There would appear to be some reason for suspecting that in the present Thurstone schedule too much weight has been given to "physique," but not enough to "consistency"; that less weight could be given to "impulsions" by comparison with "ascendance" and "drive"; that, although there seems to be agreement between the two studies that "sex" and "family" are relatively unimportant, nevertheless, they should be still further reduced to the same level as "childhood" and "sleep." The important categories on which there is agreement are "sociality," "ascendance," and "drive."

Three minor criticisms complete this article:

1. There are still in the schedule duplicate items—duplicates not only in the sense of having similar context, but also in the fact that they are almost identical in their statistical characteristics.
2. The peculiarities of the distributions of certain items (e.g., 190), combined with a subjective evaluation of their inappropriateness, suggests that there still are some items which do not properly belong in this schedule.
3. It has been suggested to the writer by Dr. R. R. Willoughby that the value of the schedule might be increased if it were completed speedily, under the stress of some urgency, induced either by the imposition of a time limit, or by requesting subjects to give only their first reactions—as they might, for example, in a free-association experiment. Implicit here is the suggestion that, in the measurement of degree of psychoneurosis, question of fact is subordinate to emotional attitude. If this be so, it is recommended that the question mark be omitted from the check-margin, and that subjects be requested to give their first reaction as either "yes" or "no," skipping an item only if quite unable to answer it.

*55 Eustis Street
Cambridge, Massachusetts*

ANNOUNCEMENT OF TRIPS TO GERMAN AND AUSTRIAN
PSYCHOLOGICAL INSTITUTES AT THE CONCLUSION OF
THE INTERNATIONAL CONGRESS OF PSYCHOLOGY
IN COPENHAGEN, AUGUST 22-27, 1932

By agreement with the German and Austrian members of the Vorstand of the International Society of Psychology, the Vorstand of the German Gesellschaft für Psychologie will arrange for foreign members to visit the larger German and Austrian psychological institutes after the Congress. The organization of the round trips is in the hands of Professor D. Katz, University of Rostock. At the earliest possible date Professor Katz will get in contact with the proper persons to carry out the suggestions of members who are interested. Additional announcements concerning these trips will be made later. For further information write to Professor K. Katz, University of Rostock, Germany.

BOOKS

TRAVIS, LEE EDWARD. *Speech Pathology: A Dynamic Neurological Treatment of Normal Speech and Speech Deviations*. New York: Appleton, 1931. Pp. xxxiv+331.

The book falls into three parts. In the first part stammering is associated with right- and left-handedness, and referred to bilateral rivalry (S. T. Orton) and to "dominant gradients" (Child, Herrick) in the cerebrum. The second part deals with organic and functional deviations in which normal habit formation is important and to which Gestalt notions of learning are applied. The third part is based on Head's *Aphasia and Kindred Disorders of Speech*.

In the first part the difference in development of the two sides and the tendency to use one side for skilled movements are naturally correlated with a "dominance" of the opposite cerebral hemisphere. The use of the right or left hand indicates this sidedness, though the use of the eyes and feet may also be considered.

Although the speech mechanism is bilateral, its function is considered unilateral and stammering is referred to the fact that a single cerebral hemisphere fails to dominate. This failure of a single side of the cortex to lead is supposed to be due either to an original indeterminate condition or to false training of the original non-dominant side.

Either the rearing of the stammerer has failed to give him a dominant side or it has violated a physiological bias and made him partly right- and partly left-sided. Therefore, great stress is laid on discovering in each case of stammering some form of ambidexterity or some indication that a left-handed child has been trained to be right-handed. This assumed rivalry between the two sides of the cerebrum is the one explanation of stammering.

The evidence presented is not convincing. In the first place, the determination of sidedness is uncertain. The use of the hand, eye, and foot is often inconsistent, and it is uncertain what significance is to be given to each in judging sidedness. The tests used are none too definite; there is no proof that facility in mirror-tracing or tendency to mirror-writing indicate handedness; and the classification of the use of two-handed implements like broom or bat into "right" and "left" is arbitrary. The testimony of relatives and friends has to be "evaluated" (p. 175). X-ray studies of the bones, like those of Menees and Holly, may furnish unequivocal criteria, but the data are not yet available. And in the second place, the data cited for the incidence of sidedness and stammering (pp. 60, 61, 139) are not consistent. J. M. Fletcher (*The Problem of Stuttering*, 19, 28) dismisses the association of handedness with stammering, and Wallin in a recent study has shown the doubtful character of such statistics.

The notion of a "dominant gradient" has one advantage over the older "center"; it implies a path of discharge which may shift, rather than a fixed clump of "controlling" cells. But, whatever explanatory value it may have had when the flow was the result of a gradient from a focus of metabolic instability, that explanatory value is quite lost when the "gradient" becomes subject to habit formation and shifts without any hint of the forces determining the "instability" on which the flow depends. It becomes a mere imaginative description of the shifting.

The incoordinations of stammering are vertical, articulation *vs.* chest movement, thorax *vs.* abdomen; they are *not* lateral as one might expect if they are to be referred to bilateral cerebral rivalry. Yet the author assumes that the thorax is dominated by one hemisphere and the abdomen by the other; that explains their action in opposite phase. But such movements in opposite phase occur in normal vigorous speech and in staccato singing; the author's kymograph tracings indicate forced and explosive utterances rather than abnormal innervation.

One can agree that J. M. Fletcher is mistaken in making stammering an emotional maladjustment; it is a just observation that such emotional disturbances figure as formative and precipitating factors rather than as fundamental causes. But his neurological speculation leads the author to ignore the close relation of stammering to tics and minor hysteria. The stammerer is obviously affected by social situations; sometimes he is able to read fluently; sometimes he can never read fluently; sometimes he can speak in public; sometimes he can never speak in public; he can sing but he cannot speak; now one group of syllables gives him trouble, now another. Stammering is strikingly subject to suggestion. It is impossible to explain all that in terms of bilateral cerebral rivalry. Fletcher and Travis are good mutual correctives; Travis' strictures on maladjustment and Fletcher's strictures on handedness are both to the point.

The treatment of stammering is evidently one of re-education, involving three months to two years and employing any means available, psychoanalysis, regimen, exercises, even "wholesome suggestion" (p. 185) with about the usual success (pp. 191-192). An unsatisfactory prescription is based on the neurological speculation; it is certainly unnecessary to retrain people of settled habits to the use of the unskilled hand and to discourage bimanual activities like piano playing and typing (walking, riding a wheel, driving a car should be included). This occasional remedy is heroic but, beyond impressing the patient, can have little to do with a cure.

In the second part, dealing with organic defects and various bad habits like lisping and persisting infantile speech, the need of careful physical and mental examinations is stressed and the suggestions for treatment are good. Mere repetition is a doubtful factor in learning, and certainly speech must be handled in wholes. But the "sound" itself, which is given as the unit, is often

fragmentary or a mere logical construction; the smallest possible unit is the syllable, and the syllable in a breath group.

The sketch of the aphasias following Head does not have much relation to the other parts of the book. Indeed, the discussion of actual lesions raises the question why obvious interference with the action of the "dominant" hemisphere does not bring stammering in its wake; the fact is that the aphasias are seldom complicated by stammering.

The publishers state that the book is intended to be a comprehensive text-book on the causes, diagnosis, and treatment of speech disorders of all kinds; it means to be both scientific and practical. The discussion of causes and symptoms is planned to be systematic and detailed. Straightforward credit is given for ideas and material and the book is fairly well documented although there is no bibliography. The index, however, is good—though J. M. Fletcher and H. Fletcher have been confused. There is an appendix containing drill material to which Chapter IV, "General Examination Methods," and the tests of pp. 215-225 might well have been relegated.

Frequent references in the text show that the book is addressed to teachers. There is a summary for the use of students. A glossary covers most of the terms; by an odd slip nystagmus is defined as "a continuous rolling of the eyeball." But the glossary does not include some of the author's own words like intellection, disconcertion, defection (lesion?), exertive, electrovital, normalcy. The style is clumsy and pedantic; the following sentence is a sample: "When there is no sufficiently dominant gradient of activity present in either hemisphere, or when such is resident but not heeded in the right hemisphere, what I conceive to be the pathophysiological subsoil of the essential speech dysfunctions obtains." In the matter of printing the book is excellent and the material is displayed for textbook use.

R. H. STETSON.

Oberlin College
Oberlin, Ohio

ACHILLES, P. S. [Ed.] *Psychology at work*. New York: Whittlesey House (McGraw-Hill), 1932. Pp. xiv+260. \$2.50.

ALLEE, W. C. *Animal aggregations: a study in general sociology*. Chicago: Univ. Chicago Press, 1931. Pp. x+431. \$5.00.

BALDUS, H. *Indianerstudien im nordöstlichen Chaco*. (*Forsch. z. Völkerpsychol. u. Soziol.*, 11.) Leipzig: Hirschfeld, 1931. Pp. x+230+viii.

BECKER, H. *Systematic sociology*. (On the basis of the *Beziehungslehre* and *Gebildelehre* of Leopold von Wiese.) New York: Wiley; London: Chapman & Hall, 1932. Pp. xxii+772. \$6.00.

BURNHAM, W. H. *The wholesome personality: a contribution to mental hygiene*. New York & London: Appleton, 1932. Pp. xxvi+713. \$2.75.

CANTIL, H. *General and specific attitudes*. *Psychol. Monog.*, 1932, 42, No. 192. Pp. x+109.

- DOLL, E. A. Twenty-five years: a memorial volume in commemoration of the twenty-fifth anniversary of the Vineland Laboratory, 1906-1931. Vineland, N. J.: Training School, Dept. of Res., 1932. Pp. xxii+135.
- DUNLAP, J. W., & KURTZ, A. K. Handbook of statistical nomographs, tables, and formulas. Yonkers, N. Y.: World Book Co., 1932. Pp. x+163.
- KULP, D. H., II. Educational sociology. New York, London, & Toronto: Longmans, Green, 1932. Pp. xvi+591. \$3.50.
- MCDUGALL, W. World chaos: the responsibility of science. New York: Covici, Friede, 1932. Pp. 119. \$1.25.
- MOORE, E. S. The development of mental health in a group of young children: an analysis of factors in purposeful activity. *Univ. Iowa Stud.: Stud. Child Welfare*, 1931, 4, No. 207. Pp. 128. Paper, \$1.15; cloth, \$1.50.
- MOSS, F. A., & HUNT, F. Foundations of abnormal psychology. New York: Prentice-Hall, 1931. Pp. xii+548.
- PRATT, C. C. The meaning of music: a study in psychological aesthetics. New York & London: McGraw-Hill, 1931. Pp. viii+253.
- PETRAK, L. A. An experimental study of pitch recognition. *Psychol. Monog.*, 1932, 42, No. 193. Pp. 124.
- ROSSMAN, J. The psychology of the inventor. Washington: Inventors Publ. Co., 1931. Pp. xii+252. \$3.00.
- SHAW, C. R., & MOORE, M. M. The natural history of a delinquent career. (*Behav. Res. Fund Monog.*) Chicago: Univ. Chicago Press, 1931. Pp. xvi+280. \$3.00.
- SWAIN, B. Fools and folly during the Middle Ages and the Renaissance. New York: Columbia Univ. Press, 1932. Pp. vi+234. \$3.00.
- SYMONDS, P. M. Diagnosing personality and conduct. New York & London: Century, 1931. Pp. xvi+602. \$4.00.
- THURNWALD, R. [Ed.] Soziologie von Heute: ein Symposium der Zeitschrift für Völkerpsychologie und Soziologie. Leipzig: Hirschfeld, 1932. Pp. viii+138.
- UMSTATTD, J. G. Student self-support at the University of Minnesota. (*Univ. Minn. Coll. Problems Ser.*) Minneapolis: Univ. Minn. Press, 1932. Pp. xiv+205. \$2.50.

RELATION OF ECONOMIC DEPRESSION TO DELINQUENCY, CRIME, AND DRUNKENNESS IN MASSACHUSETTS*†

From the Psychological Laboratories of Clark University

VERNON JONES

The relation of economic forces to crime is not ordinarily thought of, perhaps, as a psychological problem, and yet the only possible way in which such a relation could exist—if we can prove that there is a relation—is through the character, the mentality, and the emotional factors of man. The human organism is the connecting link between the stimulation of economic forces on the one hand and the responses called good character or crime on the other. If man's mental or moral or emotional powers are brought into play or affected by his being this link, then and to that extent it would appear that the study of such relationships would be a suitable field for psychological study.¹

But first of all it should be shown that there is a relation between economic forces and crime before any elaborate studies are undertaken to study the psychological processes involved. It is the purpose of this study to make a few statistical soundings to see if any relations are really apparent. It would seem that if any direct relationships exist they should show up in such a mighty swing of the economic pendulum from the prosperous days of 1928 and 1929 to the depression days beginning in 1930.

*Accepted for publication by Carl Murchison of the Editorial Board.

†I wish to acknowledge my indebtedness to State officers for official reports in the Departments of Correction, Labor and Statistics, Public Works, Corporations and Taxation, and to the Juvenile Court and Police Department officials in Worcester, Massachusetts.

¹Perhaps this word of explanation may seem unnecessary since crime is an important aspect of human behavior, and psychology should have to give no explanation for studying its genesis or nature. However, I am struck with the paucity of studies in the psychological literature on the relation of economic factors to this or any other form of behavior. Such problems have been singularly shunned as if they did *not* have interest for psychology.

ECONOMIC DEPRESSION AND THEFT

The first step in studying the connection between economic conditions and crime is to distinguish among the various types of crime. The relation might be very different for stealing and for violating the motor vehicle law. The former might increase in depression days due to need or idleness, whereas the latter might decrease due to fewer persons being financially able to operate automobiles. The trend in crimes such as assault and murder may be different from that in drunkenness. The trend in crimes among adults, many of whom are thrown out of work in days of depression, may be different from that among children, the vast majority of whom are employed with school work in days of adversity as well as in those of prosperity.

If economic conditions have any influence on amount of crime, we should expect this effect to be especially marked in connection with the various forms of theft: larceny, breaking-entering-and-larceny, stealing automobiles, and robbery. A study was therefore made of the number of arrests for these offenses for all the towns and cities of Massachusetts for each year from 1920 to 1931 inclusive. The figures given are for indices based on arrests for crimes which are ordinarily termed crimes against property. This includes mainly larceny, breaking-entering-and-larceny, the stealing of automobiles, and all forms of fraud. The number of arrests for these offenses during the year 1928² was used as the base for computing the indices. This number was 14,580. The indices for the various years are given in Table 1. A graphical representation is given in Figure 1.

Explanation of Indices. The indices are very convenient to work with and are easy to interpret. The uncorrected indices, that is, those uncorrected for growth in population, are simply obtained by dividing the figure for 1928 into the corresponding figure for each year and multiplying the result by 100. The index for a given year is, in other words, a percentage figure in terms of 1928. The year 1928 was decided upon arbitrarily as the base. Referring to Column 5 in Table 1, the index of 120 means that there were 120 per cent as many offenses committed in 1931 as in 1928. If one

²The year used here was for November 1, 1927 to November 1, 1928. All other years are for a similar period. This period rather than the calendar year was adopted because it is the one used by the Massachusetts Department of Correction.

TABLE 1
INDICES ON ARRESTS FOR THEFT AND ON EMPLOYMENT
FOR THE YEARS 1920-1931*

Year	Theft (cor- rected)	Theft including robbery (cor- rected)	Employ- ment (cor- rected)	Theft (uncor- rected)	Employment (uncor- rected)	Population
1931	115	117	74	118	76	4,289,331
1930	100.5	101	87	102.5	89	
1929	98	98	102	99	103	
1928	100	100	100	100	100	
1927	93	93	108	92	107	
1926	104	105	113	102	111	
1925	109	—	112	106	109	
1924	106	—	111	102	109	
1923	94	—	129	90	123	
1922	107	—	120	101	113	
1921	123	—	115	114	107	
1920	99	—	139	91	129	3,852,356

*The base for the indices was 1928, in which year there were 14,228 arrests for all forms of theft exclusive of robbery. When robbery is included, the number becomes 14,580. The corrected indices are corrected for growth in population. The employment figures are based on the average number of wage-earners per month for the different years in all industries in Massachusetts. The number employed in 1928 was 540,927. The main crimes included in our tabulation are: larceny including stealing of automobiles, breaking-entering-and-larceny, breaking and attempted larceny, and all types of fraud. Robbery, which is ordinarily classified as a crime against persons, is combined with the offenses in some cases; when it is, a note is made of the fact.

wished to get a percentage figure between 1931 and 1923, he would need only to divide the index of 120 by that 90; that is, there were 33 per cent more offenses in 1931 than in 1923. The corrected indices are interpreted in a similar manner. The only difference is that allowance has been made here for growth in population, so that we can say, for example, in the case of 1931 that the percentage of offenses per 100,000 in the general population of that year was 20 per cent greater than the offenses per 100,000 in 1928. In general, the corrected indices will be more valuable for our purposes, but if the reader wishes to determine the actual number of offenses committed each year, he should, of course, use the uncorrected indices in connection with the number of offenses for 1928, which is given in every case.

The most striking facts to be noted in connection with the indices

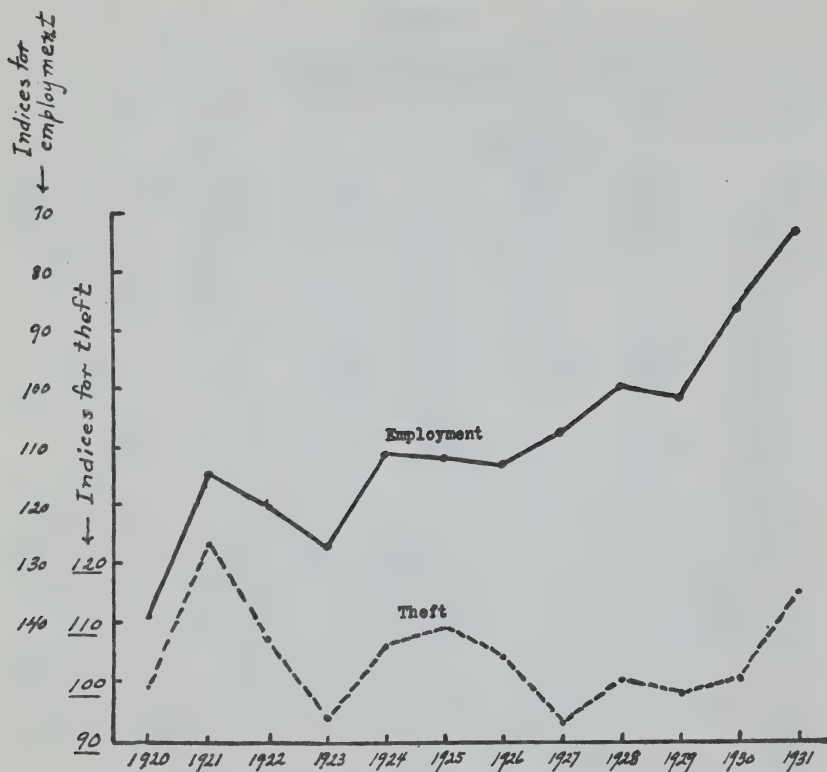


FIGURE 1

TRENDS IN THEFT AS COMPARED WITH EMPLOYMENT FOR THE YEARS 1920-1931

given in Table 1 are the high peaks in crimes of theft in the years 1931 and 1921. It is very interesting and hardly a matter of coincidence that these two years also represent the peaks in unemployment in Massachusetts during the twelve years under review.

This, however, represents only a part of the relationship which is evident in the figures. The rises and falls in the curves of unemployment and crime correspond in nine out of the eleven years. This is shown best by Figure 1. The top curve represents the indices for crimes against property, the bottom one the curve for unemployment—or rather, to be more exact, the bottom curve shows number of wage earners employed in all manufacturing industries. But since our indices have been corrected as accurately

as possible for changes in population, it is believed that the changes in the curve from one year to the next will be quite a reliable measure of relative unemployment. No claim is made that the employment indices of 1920 and 1921, for example, are strictly comparable to those of 1928, 1929, and 1930, even though corrections have been made for changes in population. It is perfectly possible that a greater percentage of the population, especially the female population, desired to work in these later years. Also there may have been changes in the percentages of persons employed in other lines of activity besides manufacturing during the range of years included. However, it is believed that the change in unemployment from one year to the next is very reliably portrayed, and it is only the rises and falls in our curves from year to year, or for a very narrow range of years, that we are interested in in this article. We wish to know if the year in which unemployment is rapidly rising over the previous year or two is accompanied by a rise in crime over the corresponding period.

If we examine the two curves in Figure 1, we find that unemployment has risen rapidly in 1921 over 1920; we note the same trend for crime. In 1922 unemployment has dropped and so has crime. The change from 1922 to 1923 is similar for both curves. The change from 1923 to 1924 is similar. The change from 1924 to 1925 is different, the unemployment slightly decreasing and crime increasing. The change from 1925 to 1926 is similar in direction. The change from 1926 to 1927 is very different; one curve is rising about as rapidly as the other is falling. The changes for all remaining years are similar. By way of summary, therefore, we can say that the directions of change in the curves of unemployment and crimes against property are similar in nine out of eleven comparisons.

In order to determine how the facts on robbery would influence the figures already reported, this offense was added to the other cases of stealing for the years for which facts on it were available. This is the only crime which could be considered stealing which is not included in the set of figures with which we have just been dealing. It is ordinarily classified among crimes against persons rather than with the other crimes of stealing, which are classified as crimes against property. From an inspection of Column 4 in Table 1, it is seen that the inclusion of robbery does not affect the indices very much except in 1931 where it increases the index

by two points. Since our totals upon which the indices are based are very large and the number of cases of robbery is, as compared with other forms of stealing, quite small, the increase in this one offense must have been very large. Upon going back to the original figures we find that the number of cases of robbery for the years 1927-1931 are as follows: 342, 352, 397, 727. The increase in 1931 over the preceding year was, therefore, 83 per cent.

If we should attempt to draw any generalized conclusion from the facts presented up to this point, it would be that there is unmistakable evidence of a tendency for drastic increases in unemployment to be paralleled by increases in crimes of theft, ranging from shop-lifting and stealing automobiles to breaking-and-entering and robbery. From whatever evidence it has been possible thus far to gather, it seems that the increases in unemployment have preceded by some amount the increases in crime—at least this appears to be true in the case of the big fluctuations. It is to be regretted that no clear-cut evidence can be given on the amount of this lag. From the trend of the graph (Figure 1) at the 1930 depression, it would appear that the lag in crime near the beginning of a depression is below what one would expect in light of the unemployment figures by several months, perhaps something like 5 to 8 months. The rise of unemployment and crime in 1921 over 1920 seems to be perfectly parallel, thus indicating no lag in this case. However, it should be stated that 1920, though a good year in comparison with other years represented in the graph, was a much poorer year for Massachusetts industries than 1919. The employment in 1920 had already dropped 18,004 from that of the preceding year, and 23,378 from that of 1918, before the sharp fall of 116,761 in 1921. Thus there is a lag between increase in unemployment and increase in crime in 1921 as well as in 1930.

It is interesting to note that Raymond Pearl (4, pp. 161-168) in his *Biology of Population Growth* sees evidence of such lag in comparing certain price indices and amounts of crime. He quotes Parmlee (3, pp. 71-86) to the effect that there is a general tendency for crime against property to increase as the price of cereals rises, and that there is a "noticeable" lag. However, neither of these writers gives any further hint as to what the size of this lag is. It seems quite probable that the amount of it will vary with localities and with different types of employment. Its significance is also a matter of some conjecture, but it is not unlikely that the

TABLE 2
INDICES FOR VIOLATIONS OF MOTOR VEHICLE LAW, FOR PASSENGER
AUTOMOBILE REGISTRATIONS, AND FOR EMPLOYMENT FOR
THE YEARS 1923-1931*

Year	Violations of motor vehicle law (corrected)	Employment (corrected)	Automobile registrations (corrected)	Violations of motor vehicle law (uncorrected)
1931	97	74	109	100
1930	121	87	111	123
1929	106	102	114	107
1928	100	100	100	100
1927	92	108	93	91
1926	74	113	93	72
1925	75	112	—	73
1924	76	111	—	73
1923	63	129	—	60

*The base used was 1928. There were 45,520 violations of the motor vehicle law that year. For explanation of the corrected indices, see the footnote to Table 1. Arrests for violations of this law were not tabulated separately before 1923.

degree to which savings hold out, the hope of getting a job, the opportunities for idling without the pinch of cold, and the general character and stamina of the individuals concerned are important factors entering into it.

ECONOMIC DEPRESSION AND VIOLATIONS OF THE MOTOR VEHICLE LAW

It has already been intimated that the relation between economic conditions and crime may be very different for different types of crime. In the last section it was shown that crimes of theft increase with the onset of depressions; in this section it will be shown that violations of the motor vehicle law³ have materially decreased in the second year of the present depression. The index for such violations decreased from 121 in 1930 to 97 in 1931. (See Table 2 and Figure 2). This tendency for motor law violations to decrease in times of depression and unemployment, if

³Violations of the motor vehicle law include all offenses committed in connection with the operation of automobiles on the highways. It does not include theft of automobiles, of course, nor does it include violations of traffic ordinances, such as parking regulations, in cities.

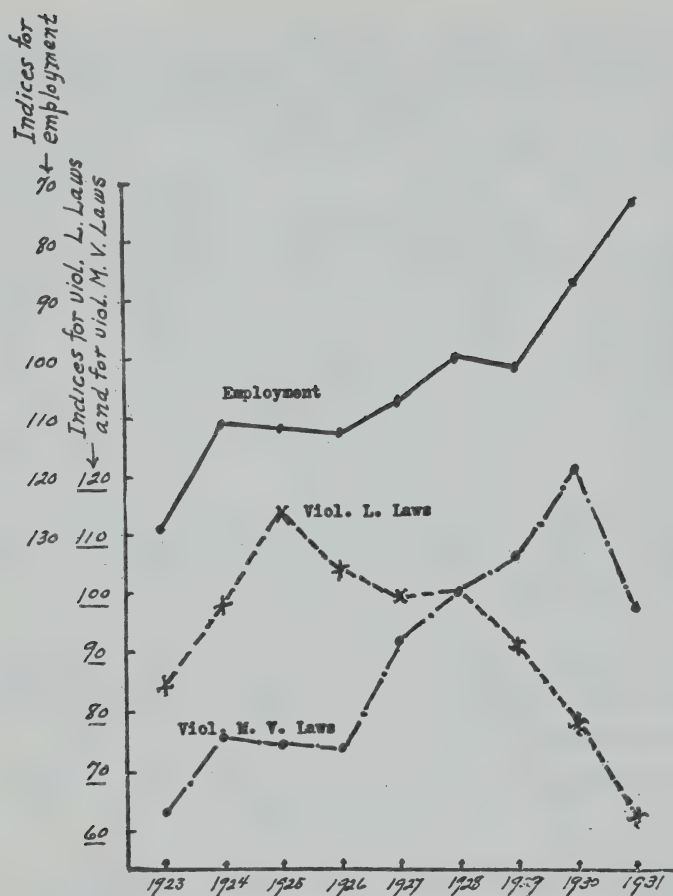


FIGURE 2

TRENDS IN VIOLATIONS OF LIQUOR LAWS AND IN VIOLATIONS OF THE MOTOR VEHICLE LAWS AS COMPARED WITH EMPLOYMENT FOR THE YEARS 1923-1931

found consistently, might be interpreted to mean that fewer people can afford to drive automobiles during such periods.

But further inspection of Table 2 will show that for 1930, the first year of the depression, the trend is in exactly the opposite direction. Here the index of employment has dropped to 87 from an index of 102 in 1929, and accompanying this has been an increase in the index of arrests for violation of the motor vehicle law from

106 in 1929 to 121 in 1930. This means that in the *first* year of the depression increase in unemployment is attended by increase in the violations of this law. And this latter trend for unemployment and violations of this law to move in the same direction is further noted in the trends in the minor fluctuations from 1923⁴ to 1930, where, in every case except one, the indices for unemployment and violations rise and fall together.

Can these apparently opposing trends be interpreted so as to fit into one logical picture? It is not certain. There are a multitude of factors influencing the violations of the motor vehicle law, and there are a large number of ways in which different degrees of economic prosperity and adversity affect the population. Besides, as has already been said, the purpose of this article is to do some prospecting for trends and relationships and not to explain how the forces work which account for these trends. However, the following possible explanation may be offered. Perhaps the relation is to an important extent a function of the number of cars on the roads and the number driven by careless and irresponsible persons. In the periods of economic prosperity preceding a depression there is an appreciable extension in the ownership of automobiles. This is shown by the rise in the index of passenger car registrations in 1929. This may bring in to the automobile-driving class a greater proportion of careless drivers. With the onset of the depression there may be an appreciable delay or lag before the automobile is sold or before the use of it is materially decreased. Indeed the use of it may increase due to the greater amount of idleness attending the increase in unemployment. Though the use of the car may increase, the expenditures for repairs and for necessary equipment, such as tires, may decrease. All of these points must be investigated before we can be sure, of course. But, if this reasoning is correct, then we would not expect the big drop in the number of cars on the roads to occur until quite a little while after the depression had got under way. This lag might well be sufficient to tide over the minor fluctuations in unemployment such as appeared in 1923 to 1930. The increase in violations in 1930 may be due partly⁵ to

⁴It is to be regretted that records cannot be obtained concerning the trends in the violations of this law in the depression of 1921.

⁵Some other factors, not directly connected with the swings in the economic conditions, which probably contribute toward the decrease are improved roads and increase in number of traffic signals and other safety measures.

this tiding-over process which keeps the cars on the roads for a while, plus the factors mentioned above of idleness and of depreciated condition of cars in the hands of those near the margin. Those who have been able to hold their cars over to the second year of the depression, that is until 1931, are smaller in number and may be a more highly selected group from the point of view of carefulness in operating their cars.⁶

If this proposed interpretation is correct, then the facts in this section on the relation of pronounced economic depression upon crime fits in well with the findings in the preceding section on the relation of economic conditions to theft. We might well expect crime against property to increase during a period of pronounced and prolonged unemployment; we should expect during the same period fewer automobiles to be on the road and hence fewer violations of the motor vehicle law.

ECONOMIC DEPRESSION AND VIOLATIONS OF THE LIQUOR LAWS

In looking over the trends in crimes in recent years in Massachusetts it was found that very appreciable changes are apparent not only in the two offenses which have been discussed thus far, but also in violations of liquor laws. This offense does not include drunkenness; it covers such things as selling, transporting, manufacture, and possession of liquor by unauthorized persons.

Upon studying the figures on this offense it immediately became apparent that a very noticeable drop in such violations had taken place since 1928, and that the rate of the decrease had become accelerated in the depression years of 1930 and 1931. Table 3 gives the indices for this offense for the years 1923 to 1931 inclusive. Facts for the years before 1923 were not available. The outstanding

⁶An attempt was made to see if there has been a drop in 1931 in the degree to which cars are run as well as a drop in the number of cars registered. This effort failed, however. The income to the State from gasoline has increased somewhat, but the rapid increase in busses and trucks, which of course consume a great deal of gasoline per mile, may more than offset the increase. The State offices do not keep separate tabulations on gasoline consumed by passenger cars. Mr. C. G. Hubbell, Statistician in the Registry of Motor Vehicles, State Department of Public Works, says in a letter to the writer: "Unquestionably the larger number of busses are using more and more gasoline every year. The reduction in violations of the law and the lower fatality figures are due, I believe, principally to business conditions which have undoubtedly driven off our roads many 'undesirables.' I refer to those young and irresponsible drivers who cause a large percentage of trouble on our highways."

TABLE 3
INDICES ON VIOLATIONS OF LIQUOR LAWS AND ON EMPLOYMENT
FOR THE YEARS 1923-1931*

Year	Violations of liquor laws (corrected)	Employment (corrected)	Violations of liquor laws (uncorrected)
1931	62	74	65
1930	78	87	80
1929	91	102	92
1928	100	100	100
1927	99	108	98
1926	104	113	102
1925	114	112	112
1924	98	111	95
1923	84	129	81

*The base used was 1928. The number of violations that year was 10,714. Arrests for this offense were not kept separately before 1923. See footnote to Table 1 for an explanation of the meaning of uncorrected indices.

facts shown by Table 3 are the rather rapid rise in violations from 1923 through 1925, the relative steadiness in numbers from 1926 through 1928, and the rapid fall beginning in 1929. These trends are shown graphically in Figure 2.

The decrease from an index of 92 in 1929 to 65 in 1931, paralleling as it does the rise in unemployment, may strike one at first as another manifestation of the influence of economic forces on crime. However, there is certainly another big factor entering in which accounts to some extent for this drop. It is the change in conditions of enforcement of the Volstead Act and other liquor laws. From December 1, 1930, the state and municipal police have been in effect prohibited by law from arresting for the manufacture and transportation of liquor. This change was brought about by repeal at the polls in November, 1930, of the so-called Baby Volstead Act. Thus a change in the minds of the law-makers and the general voting public as to what the policemen could arrest for, supplemented perhaps in many communities at least by a letting down on the enforcement of the laws which were left, has been one factor accounting for the drop in the index numbers from 1929 on. Whether or not the economic factor is another important one it is impossible to say from our data, because we have no way of estimating the percentage of the decrease which is accounted for by the enforcement factor just mentioned. It seems

quite probable, however, that during a pronounced economic depression the liquor trade would decrease, and therefore that the violation of the liquor laws would decrease. Officers having to do with the actual job of law enforcement, with whom the writer has talked, assure him that the economic factor has undoubtedly cut down on the business of "speakeasies" and has been a very important influence in bringing about the drop in 1930, and especially in 1931. But they speak on the basis of their experience and of course give no objective evidence which could be presented here.

In the study of the trends for the years 1923 to 1929 with the minor fluctuations in business we find no consistent tendency for violations of the liquor laws to go with, or to go against, the changes in unemployment. The only part of Table 3 which shows any parallel trends is that for the years from 1929 to 1932, which have just been discussed.

RELATION OF ECONOMIC DEPRESSION TO DRUNKENNESS

The trends in drunkenness and in unemployment are shown in Table 4 and Figure 3. In order to study the conditions in drunken-

TABLE 4
INDICES ON ARRESTS FOR DRUNKENNESS AND ON EMPLOYMENT*

Year	Drunkenness (corrected)	Employment (corrected)	Drunkenness (uncorrected)
1931	92	74	95
1930	89	87	91
1929	92	102	93
1928	100	100	100
1927	100	108	99
1926	101	113	99
1925	105	112	102
1924	110	111	106
1923	107	129	103
1922	99	120	93
1921	80	115	73
1920	54	139	46
1919	106	—	97
1918	126	—	114
1917	177	—	159
1916	161	—	143
1915	148	—	130
1914	152	—	133

*Base—1928. The number of arrests for drunkenness that year was 81,483. The meaning of corrected indices is explained in the footnote to Table 1.

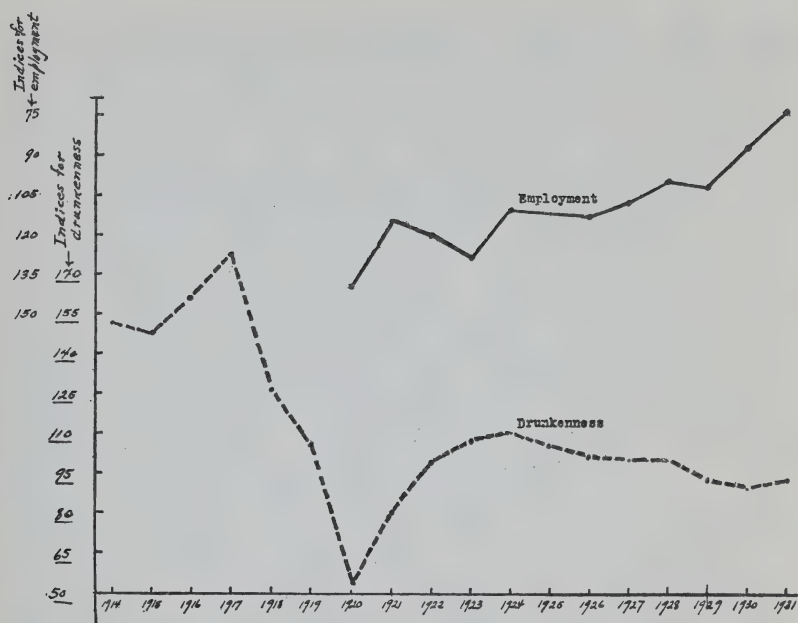


FIGURE 3

TRENDS IN DRUNKENNESS AS COMPARED WITH EMPLOYMENT

ness over several years before the national prohibition amendment went into effect, figures were obtained back to 1914. The figures on arrests for drunkenness are quite interesting. If we refer to the indices which have been corrected for growth in population, it will be seen that arrests for drunkenness in 1914-1917 were very high, averaging over one and one-half times those in 1928.⁷ The peak was reached in 1917. This was, on the basis of wage-earners employed in all industries, the third best year which Massachusetts has had since 1913, the earliest year for which we have figures. It was surpassed by 1918, which stands first, and 1919. If we lay aside all corrections for growth in population and deal with plain figures, we find the total arrests for drunkenness in the three years,

⁷Before the National Prohibition Amendment became effective in 1919, Massachusetts had had a system of local option since before 1900. Each year in each town and city the people voted for or against the licensed open saloon in their respective localities. In practice the cities usually voted wet and the small non-manufacturing towns often voted dry.

1915, 1916, and 1917, to be 352,256, whereas the total arrests in 1928, 1929, and 1930 are 231,188. In other words the arrests in years 1915-1917 exceeded those of 1928-1930 by 52.4 per cent.

Looking at the matter from the point of view of arrests for drunkenness, the years 1920 and 1921 were the real prohibition years with 1918 and 1919 acting as a good introduction to the low level in drunkenness to follow in the next two years. Between 1921 and 1929 the arrests for this misdemeanor fluctuated within a relatively narrow range. There was a steady increase in arrests for drunkenness from the very low figure of 54 in 1920 to 110 in 1924. There was a small but consistent trend of decrease in such arrests from 1924 through 1930, if corrections are made for increases in population. However, if we consult the actual figures without allowing for growth in population there are two interruptions to this trend; these appear in 1927 and 1928. In 1931 there was an increase of 3 per cent over 1930 if allowances are made for growth in population, or a 4 per cent increase if we use the raw figures.

In studying these trends it is interesting to compare them with the recent findings of Frederick W. Brown (1) with regard to the trends in alcoholic mental disease and deaths from cirrhosis of the liver resulting from excessive alcoholism. He shows that there was a very steady decline in the percentage of new cases of alcoholic mental disease and of intemperate users of alcohol among all new cases admitted to civil state hospitals of New York from 1911 to 1915. This was followed by a noticeable increase in 1917, and this, in turn, was followed by a steady drop to 1920. After 1920 he finds a steady increase up to 1927, and since that time he reports the trend as irregular. He checked this trend in New York State against that in 56 hospitals in 25 states for the same years and found full substantiation for his conclusion based on the one state. His study of deaths from cirrhosis of the liver caused by excessive alcoholism shows quite similar trends to 1920, but there is no evidence of substantial increase since that time.

It will be noted that his finding on percentage of new cases of mental disease and of intemperate users of alcohol among new entrants to state hospitals agrees very closely with our findings. The only differences in the two reports of trends are (1) that he reports a steady increase up to 1927, while, according to our figures, this increase stopped in 1924; and (2) that he refers to the trend from 1927 to 1930 as irregular, whereas our figures indicate that

it is predominately downward, especially if we allow for growth in population. But the clearest trends correspond exactly in the two studies. It is interesting to note what he says about the trend in 1917 and in the year immediately following:

"There was a noticeable increase in 1917, with a steady drop to 1920. During these three years (1918-20) the prevailing public opinion favoring the conservation of food stuffs and the prevention of waste and excesses of all kinds was, it is believed, largely responsible for the success of prohibition enforcement. This may be called the real prohibition period."

So much for the trends within the drunkenness column itself. How are these trends on drunkenness related to those in unemployment? From what has already been intimated about the powerful influences which affect attitudes toward drunkenness it is apparent that it will be impossible to cull out the effect of economic forces alone. The public opinion toward waste in times of war, the educational emphasis upon decency and respectability which was persistent and strong at least before we "settled the matter" by legislation, the passing of the state laws, and finally the national amendment, which apparently were taken more seriously at first than later, the gradual change in public opinion with regard to violators of the liquor laws—all of these, we believe, have been weighty forces in connection with the arrests for drunkenness. But these have not changed suddenly and, if rather abrupt changes in economic conditions appear, then their effect, if it is pronounced, should show up in comparing adjacent years which differ appreciably in unemployment. Table 4 enables us to make such a comparison all along the line from 1920. In studying the indices for drunkenness and unemployment we find no consistent relationship. In six cases drunkenness moves in the same direction with unemployment; in four it moves in the opposite direction. In the depression of 1921 drunkenness strongly increased with increased unemployment. (See Figure 3.) In 1930 it decreased slightly with increased unemployment. In 1931 it increased a little with further increases in unemployment. In 1924, when there was a milder depression, it increased slightly with an increase in unemployment. Thus there are three cases where increased drunkenness went with increased unemployment, and one case where the opposite relation holds. This relation is, therefore, not clear enough to justify

any definite conclusion, but the weight of the evidence tends in the direction of indicating that increases in unemployment are attended by slight increases in arrests for drunkenness.

It may seem at first difficult to reconcile this probable trend with the conclusion reached earlier that violations of the liquor laws have decreased during unemployment. This could be explained on the basis of the sampling of the ordinary consumers of alcohol who are arrested for drunkenness. Brown (1) says that 65.6 per cent of alcoholic mental patients (who are the worst of those arrested for drunkenness) are on the margin between self-support and dependency, and an additional 29 per cent are dependent. This is probably not far from an accurate picture of the sampling of the population which is arrested for drunkenness. These unfortunates are among the first to lose their jobs, if they have them, in a depression, and it is quite conceivable that their drunkenness increases with their idleness as long as their money, and perhaps that of their drinking friends, holds out. The consumption of alcohol by moderate drinkers, however, may be appreciably decreased in times of unemployment.

RELATION OF ECONOMIC DEPRESSION TO JUVENILE DELINQUENCY

When juvenile delinquency is separated out for study, it is found to be no inappreciable part of the problem of crime. Facts were not easily available by which adult and juvenile *arrests* could be distinguished, but it was found that separate statistics of a detailed nature could be obtained concerning the *number of cases begun* in the municipal and district courts against juveniles under 18 years of age. It is not claimed, of course, that the figures for this category would be the same as those for arrests, but it is believed⁸ that these figures will represent accurately the trend in the more serious cases of delinquency for adjacent years. Table 5 gives the indices for years 1920-1931 inclusive for all offenses combined. The totals on all offenses are used because a study of the separate offenses of children seemed to show no reliable evidence of specialized trends such as was found in the case of violations of the motor vehicle law and the liquor law in our analysis of total arrests of adults and children.

Over half of the offenses of the delinquent group are some form

⁸Some of our indices based on arrests for all offenders have been compared with figures on "cases begun," and the same trends are shown.

TABLE 5
INDICES FOR JUVENILE DELINQUENCY AND FOR EMPLOYMENT
FOR THE YEARS 1920-1931*

Year	Delinquency (corrected)	Employment (corrected)	Delinquency (uncorrected)
1931	93	74	96
1930	99	87	101
1929	101	102	102
1928	100	100	100
1927	92	108	91
1926	86	113	84
1925	97	112	94
1924	91	111	87
1923	82	129	78
1922	84	120	78
1921	108	115	101
1920	102	139	94

*Base—1928. The number of offenders brought before the court was used as the basis for measuring delinquency. The figure for 1928 was 6747. Figures on the arrests for children were not available for the entire state, but from the figures available it seems clear that the trend for the figures for "brought before the court" closely parallel that on juvenile arrests. The former figures will be, in general, smaller than the latter, but ordinarily will not differ from it more than 12 per cent. See footnote to Table 1 for an explanation of the meaning of uncorrected indices.

of theft, and any changes in total number of offenses from year to year are mainly reflections of changes in this crime. Of all offenders (boys and girls) coming before the Boston Juvenile Court in 1930, there were 57 per cent who were charged with larceny, or breaking-entering-and-larceny, or breaking and attempted larceny, or robbery. The corresponding figure for the Worcester Juvenile Court was 56 per cent ($N=586$). If boys alone are studied this percentage rises appreciably. In the study of 162 offenses committed by 46 boys⁹ over several years in Worcester, it was found that 74.6 per cent of the offenses fell under the forms of stealing just mentioned.

With these percentages in mind concerning the part which theft plays in the total figures, we may refer to Table 5 for a general

⁹This group which was studied somewhat intensively was small though it seems to be quite a typical one. The mean intelligence quotient was 83.1 (median = 82.0) as compared with a median of 83.5 derived from the results of 11 different studies summarized by Pintner (5, pp. 378-379). The median age of our group was 14 years and 4 months, whereas Healy and Bronner (2, p. 92) report the median age of their large sampling of 2830 delinquent boys of Boston and Chicago as being "between 14 and 15."

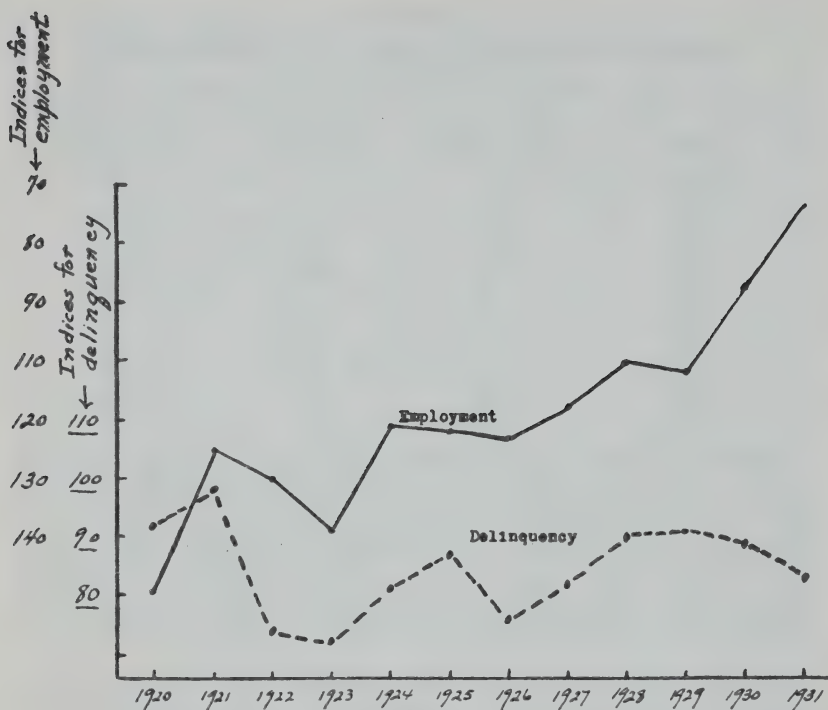


FIGURE 4

TRENDS IN JUVENILE DELINQUENCY AS COMPARED WITH EMPLOYMENT
FOR THE YEARS 1920-1931

summary of changes from year to year and for a comparison between the indices for delinquency and for unemployment. It will be noted that the changes have been irregular. No definite trend is observable. The average of the indices for the first three years in Table 5 is the same as for the last three years. The largest fluctuation for adjacent years is the decrease in delinquency from the high point of 108 in 1921 to 84 in 1922.

The comparison between the figures for delinquency and those for unemployment should now receive attention. Since it has been shown that these figures on total delinquencies mirror to such an important extent the trend in stealing, and since it was shown in the first section of this article that arrests for stealing (for adults and children combined) have varied quite consistently with

unemployment trends, we would expect juvenile delinquency to increase as unemployment increases. We do not find this consistently, however. In the years 1920-1922, which included the depression of 1921, this tendency is quite apparent, but in the years 1929-1931 the opposite trend is seen. If we analyze further the rises and falls of delinquency and unemployment year by year for the twelve-year period, we find that from 1920 through 1928 there was a parallel change in delinquency and unemployment in 7 out of 8 comparisons, the one case to the contrary being for the year 1925, where unemployment decreased a very small amount but where delinquency increased appreciably. Since 1929 unemployment has increased and delinquency has decreased. The writer has no positive knowledge as to what accounts for this change in trend since 1929. Probably better methods of dealing with the juvenile delinquency problem have been adopted by judges, heads of delinquent institutions, and parole officers. It is known that there have been changes from time to time in different communities with regard to the percentage of juvenile offenders who are handled out of court. But no statistics on out-of-court cases are available for the entire state, and therefore it is impossible to allow for any influence which such changes may have had. It is believed that it is small. It may be said with certainty, however, that there has been no unanimous agreement among judges of juvenile courts in the state to encourage the handling of a larger percentage of the cases out of court. In some communities there has been the opposite tendency.

The only definite conclusion which we can draw from our data is that there is no consistent trend for increased unemployment to be attended by increased delinquency among juveniles. But there is such a trend among adults.¹⁰ This failure to find the same relationship between depression and crime for children and for adults may possibly mean one or more of the following things: (1) that

¹⁰This is shown by the fact that since stealing among juveniles has not increased consistently with increased unemployment, but has increased for adults plus juveniles, the increase must be due to a trend existing among the adults. Moreover, it is shown by the fact that the number of arrests of all persons for *all* crimes except violations of liquor laws and the motor vehicle law—almost exclusively adult crimes—has risen and fallen in perfect agreement with the unemployment figures over all the years for which we have studied this trend, viz., 1927 to 1931 inclusive. But in the last three of these years total offenses of the juveniles alone have gone exactly counter to this trend. The parallel trend must exist, therefore, among the adults.

economic forces make themselves felt more directly and more keenly among adults than among children; (2) that the schools exercise a stabilizing influence on children; (3) that the increased activities of social service agencies have a more constructive effect upon children than upon adults; and (4) that the newer corrective measures being tried by juvenile courts, delinquent schools, and parole officers are acting as more effective deterrents to crimes among children than are the legal and penal methods being applied to adults, who, after all, are probably not very amenable to improvement by any penal methods.

SUMMARY AND CONCLUSIONS

In this statistical study an attempt has been made to uncover as many significant relationships as possible between big swings in economic conditions and crime in one representative state. Special attention has been given (1) to the relation between unemployment and frequency of arrests of all offenders for crimes of stealing, for violations of the motor vehicle law, for violations of liquor laws, and for drunkenness; and (2) to the relation between unemployment and juvenile delinquency. The general conclusions reached are as follows:

1. All crimes of theft among adults, that is, persons 18 years of age or older, have increased materially during the depression beginning in 1930. They also increased in the depression of 1921. Furthermore, by making year-to-year comparisons it was found in 9 cases out of 11 that when unemployment increased, theft increased; and when unemployment decreased, theft decreased. (For the basis of this conclusion see not only Table but also p. 277.)

2. There is evidence of a lag between the onset of unemployment and the increase in crimes of theft and robbery.

3. Violations of the motor vehicle law have in general decreased with unemployment. This rule applies with only one exception from 1923 to 1930, and it applied in 1931. Presumably this relation is to some extent tied up with the number of cars on the roads. However, there was a marked increase in violations in the depression year of 1930, when we would have expected fewer cars on the roads. Perhaps this exception was due largely to the lag phenomenon; that is, with the onset of the depression there may have been an appreciable delay or lag before the money and cars owned by many in the prosperous days of 1929 were disposed of. In the

initial days of unemployment the cars may actually have been used more (but kept up less well) than in 1929.

4. Arrests for violations of the liquor laws have decreased very materially since 1929. The arrests in 1931 were only 65 per cent of those in 1928. If growth in population is allowed for, there were only 62 per cent as many arrests per 100,000 inhabitants in 1931 as in 1928. This drop may be somewhat connected with the economic depression in that perhaps some of the previous customers of the liquor law violators had no money with which to purchase liquor during unemployment. However, there is another factor which has certainly had a large part in accounting for this decrease. It is the changed attitude of the public and changes in the state enforcement laws. Since we do not know what percentage of the drop in the arrests is due to this powerful factor, we have been unable to arrive at anything definite concerning the relation of economic forces to violations of the liquor laws.

5. Trends in drunkenness from 1914 to 1931 were studied. The arrests for drunkenness in the three years 1915-1917 were a little over one and one-half times those in the three years 1928-1930. The total arrests in the former period were 352,256; those in the latter were 231,188. Judging by the number arrested for drunkenness, the years 1920 and 1921 were the years of real prohibition in Massachusetts, the years 1918 and 1919 acting as a good introduction to a relatively low level in drunkenness which was reached during these two best years. The arrests for drunkenness in 1920 were 25 per cent lower than in any other year of the 18 which we have studied, i.e., 1914-1931. There were only 46 per cent as many arrests in 1920 as in 1928; if allowance is made for growth in population, there were only 54 per cent as many as in 1928.

6. There was no conclusive evidence found of a relation between economic conditions and drunkenness. We note a pronounced tendency for increased unemployment to be attended by increased drunkenness in the depression of 1921, but in 1930 we see a mild trend in the opposite direction. The changed attitude toward prohibition in the depression period of 1930 from that of the depression period of 1921 makes it difficult to interpret our data.

7. Juvenile delinquency has decreased during the present economic depression. Allowing for changes in population, the delinquency in 1930 and 1931 combined was 4 per cent below that of 1928 and 1929 combined. In the years 1920 through 1928

delinquency rose and fell with unemployment in 7 out of 8 years, but this trend has not been maintained since that time. In each of the years 1930 and 1931 there has been a drop from the figures of the preceding year. This decrease in delinquency at a time when total adult crimes—except violations of motor vehicle laws, violations of liquor laws, and drunkenness—are increasing is striking. The failure of juvenile offenses to *rise* with unemployment, as comparable adult offenses do, might mean that economic forces make themselves felt more directly and keenly with adults than with children. But the actual decrease must mean that some constructive forces are at work. It may mean that the increased attention to pupil adjustment and character training in the schools, the redoubled efforts of social service agencies working with children, and the attempted improvements in methods of treatment of juvenile offenders in courts and corrective agencies are yielding perceptible results in the form of decreased delinquency.

REFERENCES

1. BROWN, F. W. Abstract: Alcoholic mental disease before and after prohibition. *J. Amer. Statis. Asso. Suppl.*, 1932, 27, 175-178.
2. HEALY, W., & BRONNER, A. F. Delinquents and criminals—their making and unmaking. New York: Macmillan, 1926. Pp. x + 317.
3. PARMLLEE, M. Criminology. New York: Macmillan, 1918.
4. PEARL, R. Biology of population growth. New York: Knopf, 1925.
5. PINTNER, R. Intelligence testing. (Rev. ed.). New York: Holt, 1931. Pp. xii+555.

Clark University
Worcester, Massachusetts

LA RELATION DE LA CRISE ÉCONOMIQUE AU DÉLIT

(Résumé)

Dans cette étude statistique on a essayé de découvrir toutes les relations significantes possibles entre les grands changements des conditions économiques et le délit dans un Etat représentatif. Les conclusions les plus importantes tirées sont les suivantes: (1) Tous les crimes de vol parmi les adultes ont beaucoup augmenté pendant la crise qui a commencé en 1930. En faisant des comparaisons an par an de 1920 à 1931, on a trouvé en 9 cas sur 11 que lorsque le chômage a augmenté, le vol a augmenté; et lorsque le chômage a diminué, le vol a diminué. (2) Il se montre un retard entre le commencement du chômage et l'augmentation des crimes de vol et de larcin par les adultes. (3) Les violations des lois des automobiles ont en général diminué avec le chômage. (4) Les arrêts pour les violations des

lois des boissons alcoôliques ont beaucoup diminué depuis 1929. Cependant, puisque l'attitude changée du public à l'égard des lois boissons a tant influé sur le nombre des arrêts, on n'a pu arriver à une conclusion définitive sur la relation des forces économiques aux violations de telles lois. (5) On a étudié des tendances des arrêts pour l'ivresse de 1914 à 1931. Les arrêts pour l'ivresse dans les trois ans de 1915-1917 avant la loi de prohibition ont été un peu plus d'une fois et demie ceux des trois ans 1928-1930. Les deux premiers ans de la prohibition nationale, c'est-à-dire 1920 et 1921, ont été les ans de la vraie prohibition en Massachusetts, les ans 1918 et 1919 faisant le rôle d'une bonne introduction au niveau peu élevé de l'ivresse atteint dans ces deux meilleurs ans. On n'a trouvé aucune évidence conclusive d'une relation entre les conditions économiques et l'ivresse. (7) Le délit juvénile a augmenté pendant la crise économique actuelle. Dans les ans 1920-1928 les crimes ont augmenté et ont diminué avec le chômage dans 7 sur 8 ans, mais cette tendance ne s'est pas montrée depuis ce temps. Dans chacun des ans 1930 et 1931 il s'est montré un abaissement des chiffres de l'année précédente. Cette diminution du délit à un temps quand les crimes comparables par les adultes augmentent est frappante. Elle signifie probablement que la plus grande attention aux meilleures méthodes de correction, à l'ajustement des enfants, et à l'entraînement du caractère donnent des résultats perceptibles.

JONES

DIE BEZIEHUNG DER ÖKONOMISCHEN DEPRESSION ZUM VERBRECHERTUM (DELINQUENCY)

(Referat)

In dieser statistischen Untersuchung ist der Versuch gemacht worden, möglichst viele bedeutsame Beziehungen zu entdecken, zwischen grossen Schwankungen ökonomischen Zustände und dem Betrag an Verbrechen in einem typischen Staat. Die wesentlichen Befunde sind folgende: (1) Alle Diebstahlverbrechen unter Erwachsenen haben im Lauf der Depression, von 1930 fort, bedeutend zugenommen. Vergleicht man Jahr für Jahr die Jahre 1920 bis 1930, so findet man, bei 9 Fällen aus 11, dass mit Zunahme der Arbeitslosigkeit der Diebstahl zunahm, während mit Abnahme der Arbeitslosigkeit auch der Diebstahl abnahm. (2) Man fand, dass die Zunahme des Diebstahls und des Raubes bei Erwachsenen nach dem Anfang der Arbeitslosigkeit eine etwas verzögerte war. (3) Übertretungen der Gesetze in Bezug auf Automobile haben im Allgemeinen mit der Arbeitslosigkeit abgenommen. (4) Verhaftungen wegen Übertretung der Trinkgesetze (liquor laws) haben seit 1929 ganz sichtbar abgenommen. Da aber die veränderte Einstellung des Publikums den Trinkgesetzen gegenüber die Zahl der Verhaftungen sehr stark beeinflusst hat, ist es uns unmöglich gewesen, zu irgend einer bestimmten Meinung zu gelangen über die Beziehung ökonomischer Kräfte zu Übertretungen solcher Gesetze. (5) Es sind die Tendenzen der Verhaftungen wegen Trunkenheit von 1915-1931 untersucht worden. Die Zahl der Verhaftungen wegen Trunkenheit war in den drei Vorprohibitionsjahren 1915-1917 etwas mehr als ein-und-einhalb Mal so gross wie in den drei Jahren 1928-1930. Die zwei ersten Jahre der nationalen Prohibition, d.h., 1920 und 1921, waren die Jahre wirklicher Prohibition im Staat Massachusetts, und die Jahre 1918 und 1919 dienten als eine gute Einleitung zum niedrigen Niveau der Trunken-

heit das in diesen zwei besten Jahren erzielt wurde. (6) Man fand keinen überzeugenden Beweis einer Beziehung zwischen ökonomischen Umständen und Trunkenheit. (7) Verbrechen Jugendlicher (juvenile delinquency) haben während der gegenwärtigen ökonomischen Depression abgenommen. In den Jahren 1920 bis 1928 stieg und fiel die Zahl der jugendlichen Verbrechen mit der Arbeitslosigkeit in 7 aus 8 Jahren, aber diese Richtung ist seitdem nicht beibehalten worden. In jedem der beiden Jahre 1930 und 1931 waren die Zahlen jedesmal kleiner, als im vorhergehenden Jahr. Diese Abnahme in der Zahl der jugendlichen Verbrechen zu einer Zeit, zu der vergleichbare Verbrechen Erwachsener zunehmen, ist auffallend. Sie will wahrscheinlich heissen, dass die verstärkte Aufmerksamkeit auf verbesserte Verbesserungsmethoden, auf die Anpassung des Kindes, und auf die Erziehung des Charakters bemerkbare Resultate liefert.

JONES

THE INFLUENCE OF INDIVIDUAL OPINION ON CONSTRUCTION OF AN ATTITUDE SCALE*

From the Psychological Laboratories of the University of Florida

E. D. HINCKLEY

INTRODUCTION

Until recently, all attempts to measure attitudes have taken the form of questionnaires or rating scales. Although in many cases such methods are sufficiently accurate for the purpose of the study, they do not present us with a refined technique for measurement.

In the publication, "Measurement of Opinion" (6), Thurstone criticizes all previous studies as being contributions of merely bar-diagram representations, which in no case produce a mathematically sound and scientifically accurate scale for measurement. None of the studies has presented us with a yardstick on which the unit of measurement remains the same throughout the scale. Thurstone offers a new method of attack in his articles, "Attitudes Can Be Measured" (3) and "Theory of Attitude Measurement" (8), basing his technique on mathematical laws which he develops in the additional publications: "Psychophysical Analysis" (5), "Law of Comparative Judgment" (4), "Three Psychophysical Laws" (9), "Unit of Measurement in Educational Scales" (10), and "The Phi Gamma Hypothesis" (7). On the basis of this technique, three scales have been constructed under Thurstone's supervision:

Attitude toward the Church, by Chave and Thurstone (11),

Attitude toward Prohibition, by Smith (2),

Attitude toward Militarism-Pacifism, by Droba (1).

These scales have already been used to measure differences in opinion between two groups, and also changes in opinion in the same group after certain influences were brought to bear.

In the case of each scale which has thus far been constructed according to Thurstone's technique, it has been assumed that the scale is not influenced by the attitude of the subjects used in the

*Accepted for publication by L. L. Thurstone of the Editorial Board and received in the Editorial Office, July 3, 1931.

construction of the scale. It is the aim of the present research to make a thorough experimental test of this assumption. Stated more definitely, our problem is *to construct a scale for measuring attitude toward the Negro, and to test the extent to which this scale is influenced by the opinions of the subjects used in the construction of the scale.* (By opinions is here meant opinions as to the social position of the Negro.)

It is obvious that a measuring device must not be seriously affected in its measuring function by the objects which it measures. If an inch were of one size in measuring a desk and of another size in measuring a rug, it would not be a valid unit of measurement. Likewise, an instrument devised for the measurement of a certain attitude must not be affected in its measuring function by the individuals whose attitudes are being measured. To the extent that such a scale is so influenced, its validity is impaired. The completed scale must not only transcend the groups and individuals to be measured, but also the groups and individuals used in the construction of the scale. This premise demands that the scale values of all statements be independent of the attitudes of the individuals or groups who aided in the sorting of the statements.

Let us see how such an influence could be exerted. Each subject in the construction of the scale must judge as to the degree in which each statement reflects the attitude variable. Now, his judgment might possibly be influenced by his own attitude on the question. If this be true of very many of the judges who aid in the construction of the scale, the resulting scale is not a valid measuring device. The discrimination of the value of any two statements at any portion of the scale must be the same for a person whose attitude is toward one extreme of the scale as for one whose attitude is toward the other extreme, or for one who is neutral in respect to the question. For instance, in a scale of attitude toward the church, should about 85% of the antagonistic judges declare statement "a" less antagonistic in attitude toward the church than statement "b," then it would be expected that about 85% of the friendly judges would make the same classification. Furthermore, if a scale for measuring attitude toward militarism-pacifism were constructed using militaristic subjects, it should be possible to measure pacifistic attitude just as accurately as a similar scale constructed by pacifistic judges. The present research is a study of this factor, that is, a study of the manner in which the position of the sorters on an attitude scale affects the validity of that scale.

PROCEDURE

The present research consists in the calculation of three independent sets of scale values for certain statements of opinion about the Negro: the first on a group of white southern subjects who were prejudiced against the Negro; the second, on a group of white northern subjects who were prejudiced in favor of the Negro; and the third, on a group of Negro subjects. Each set of scale values was computed in accordance with Thurstone's method, and the three sets were carefully compared.

The particular attitude variable which was chosen to be measured is that of the Negro's social rights. This particular attitude variable was chosen because it was sufficiently limited to be measurable and sufficiently debatable to furnish a large range of opinions. Furthermore, groups of subjects were available for the writer's use at the University of Florida and the University of Chicago, where differences of opinion on this attitude variable were sure to occur.

Our first real task was the selection of the statements of opinion as to the social position of the Negro, which opinions would range from the feeling that the Negro should be treated socially on a par with the animal to the attitude claiming social equality with the white man. The statements of opinion were collected from two sources. In the first place, all available literature of the Negro in the libraries of the University of Chicago and the University of Florida was scanned for brief, concise opinions. In the second place, I asked each of 200 students in my psychology classes at the University of Florida, and each of 25 college professors to write his opinion as to the social position of the Negro in as concise and clear a statement as possible. From these two sources were culled 230 statements, ranging between the two extremes of opinion stated before.

It is necessary that the statements which are used in the final scale be clear, concise, unambiguous, and that they contain no irrelevant elements. Consequently, this list of opinions was scrupulously refined according to these criteria. If there was the slightest suspicion of irrelevant elements, the statement was canceled from the list. In this way, the original list was reduced to 150 statements. Copies of this list were then given to Professors Thurstone and Kingsbury, of the Faculty of the University of Chicago, and to two graduate students, all of whom are thoroughly

familiar with the problem, with the request to check those statements which they felt were ambiguous or contained irrelevant elements. We were especially anxious that a sufficient number of unambiguous neutral statements should remain in the list so that our scale would not be split into the two extremes by the breaking down of the middle range. After all of these things had been taken into consideration, there remained 114 statements, and these make up the final list which was used in the experiment.

The care which was taken in this refining process cannot be overemphasized. The search for statements and the reduction of the original list to 114 statements extended over a period of a year. The efficiency of the execution of this process is mirrored in the uniformity and the relevance of the completed scale.

These 114 statements were then mimeographed on small cards, one statement to a card, for purposes of sorting. Five hundred envelopes were then made up, one envelope for each subject. In each envelope were placed the 114 statement cards, 11 pile-heading cards with the Roman numerals I, II,—XI, one card for the name and classification of the subject, 11 paper clips, and a copy of the directions. Later, more cards were mimeographed, and four hundred more envelopes were filled in the same way.

Each subject was given an envelope and instructed to sort the 114 statements into eleven piles according to the attitudes as to the social position of the Negro which the statements expressed. They were to put in Pile I those statements which place the Negro in the lowest social position, in Pile XI those statements which place the Negro in the highest social position, and in each of the other nine piles those statements which seemed to express that particular step in the social scale. Before starting the task of sorting the statements, each subject was requested to read over carefully the six statements given at the bottom of the instruction sheet, and to check those of the six which he could endorse as expressing most nearly his own attitude toward the Negro. The six statements are as follows:

1. Inherited qualities have predestined the Negro to the servant class of society.
2. The white and colored races should enjoy the same privileges and protection as set forth by law.
3. The Negro should be used to produce the white man's needs.
4. I do not know enough about the social possibilities of the Negro

to pass judgment upon him.

5. Our refusal to accept the Negro is not based on any fact in nature but rather on prejudice, and should be overcome.
6. So great is the social range between the highly educated Negro and the 'nigger,' that the race as a whole cannot be assigned to any one notch in the social scale.

It is clear that those subjects who check the second or fifth of these six statements have an attitude which favors the Negro, and that those who check the first or third hold an attitude which does not favor the Negro, while those who check the fourth or sixth have an attitude of neutrality on the question. The subjects can thus be divided into three groups according to their own attitude toward the social position of the Negro. In addition to the written instructions, the task was explained orally to each subject.

Eight hundred and fifty subjects were recruited from nine schools. Following is a list of the colleges which participated in the experiment:

University of Chicago	Chicago, Illinois
University of Florida	Gainesville, Fla.
Virginia Union University	Richmond, Va.
Shaw University	Raleigh, N. C.
Morris Brown University	Atlanta, Ga.
Morgan College	Baltimore, Md.
Lincoln High School	Gainesville, Fla.
Lincoln University	Lincoln Univ., Pa.
Benedict College	Columbia, S. C.

Of these 850 subjects, 600 were white students and 250 were Negroes. Practically all of these subjects were enrolled in some class in psychology, sociology, or education, in which the whole class hour was given over to the task of sorting the 114 statements. Thus practically all of the sorting was personally supervised, either by the writer or by someone who was thoroughly familiar with the problem. Every student was required to stay in the classroom until the end of the hour, even though he had finished the sorting. The writer feels that much more careful sorting was achieved by this method than if the students had sorted the statements outside of class without supervision.

It is not to be expected with such a large number of subjects that

every subject would understand the directions thoroughly, or that all of those who did understand them would follow them diligently. One tendency which revealed itself in the sortings of some subjects was the bunching of statements in one or more piles to the apparent detriment of the other piles. This phenomenon of bunching at the extremes was noticed in the case of certain of the white subjects, but was especially noticeable in the Negro subjects. Since the 114 statements are distributed with fair uniformity over the entire scale, marked bunching is a sign of careless sorting. If more than a fourth of the statements are assigned to any one pile, it will leave less than three-fourths to distribute over the remaining ten piles. Furthermore, the individual who sorts the statements in this fashion often ignores some of the piles completely. On the assumption that this bunching was due to poor discrimination and carelessness, every case having 30 or more statements in any one pile was automatically eliminated from consideration, and the results were not recorded. Several other cases were found which clearly demonstrated that the statements had been sorted with no criterion at all, or according to some attitude variable other than the social position of the Negro. These cases were also eliminated from consideration.

The envelopes of the Negro subjects were all placed together for separate consideration, and will be discussed later. The envelopes of the white subjects were divided into three groups, according to the statement on the instruction sheet which had been checked. Group I consists of those who declared themselves against the Negro by checking the first or third statement on the instruction sheet. Group II consists of those who placed the Negro high in the social scale by checking the second or fifth statement on the instruction sheet. In a third group were placed those who expressed an attitude of neutrality by checking the fourth or sixth statement on the instruction sheet. Since this neutral group could have no place in the present research, the envelopes of this group were placed with those others which were eliminated from consideration, and the results were not recorded. A few of the subjects displayed carelessness by checking one friendly and one antagonistic statement, the first and fifth, for example. Their envelopes were likewise eliminated from consideration.

The arrangement of the statements in each envelope in Group I was then recorded according to the following form:

Statement	Pile										
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
.											
.											
.											
114											

When all the envelopes in Group I had been recorded, we thus had the frequency with which each statement had been placed in each pile. These data were then rearranged in the form of a separate frequency table for each statement. The arrangement of the statements in each envelope in Group II was then recorded in the same manner, and a frequency table was made for each statement, as before, showing the frequency with which that statement was placed in each pile by the sorters of Group II.

For each statement, there were then plotted the values in the percentile column, with a smooth percentile curve drawn through them. Since this was done for Group I and Group II separately, there are two curves for each statement. For purpose of comparison, the two curves for each statement were drawn on the same coordinate axes. In order to ascertain the proper position of each statement in the scale and its relative degree of ambiguity, the scale value and quartile deviation were calculated from each curve. The scale value is the abscissa of the point where the curve crosses the 50% line. The quartile deviation is half the horizontal distance between the point where the curve crosses the 25% line and the point where the curve crosses the 75% line. The Pearson coefficient of correlation between the series of scale values in Group I and the scale values in Group II was then computed.

Although the purpose of the present investigation could be achieved by a comparison of the results of Group I with the results of Group II, the writer thought that it would be interesting to build up a third series of scale values, using Negroes as subjects. Through the cooperation of certain colleges and universities, named on page 287, it was possible to obtain several competent groups of Negro judges. It was found that most of the colleges offered courses in "Race Relations," "Racial Attitudes," or some related study. For the sake of uniformity, it was considered best to have the sorting supervised, as it was in the case of the other two groups. In excellent and fullest cooperation, the professors in charge gave a part of one class period to preliminary directions, and donated an entire class period to the sorting.

Since the attitude variable is a national as well as a racial attitude, it was deemed advisable to select judges as diverse geographically as expedient, and in this way to lessen possibility of any sectional attitude causing a bias in the scale. It will be observed that seven states are represented, extending as far north as Pennsylvania and Maryland. Although the record of each school was kept separately, there was not a sufficient difference in the distributions to provoke comment.

Reports from all schools but one indicate that the entire class concerned seemed to display a purely scientific interest in the experiment, evidently free from emotional prejudice. Several professors commented on the contributions of the opinions toward later class discussions. The one exception mentioned reported considerable resentment or prejudice toward the investigation. It appears this feeling was engendered between the previous class period and period of the sorting, and that the students concerned had had no opportunity for contact with the actual opinions contained in the individual envelopes. The height of this emotion was to such an extent that about half the class declined to act in the capacity of judges, while the other half consented. Very special attention was given to the comparison of the results of this remaining half, with the results of the other schools, but no indication of unproportional divergence could be detected.

It is not a digression here to note that the few cases admitted from the one high school were entirely satisfactory. It was thought that immaturity might have a different influence on the results in comparison with college students. All girls and boys of the high school in question were above 18 years of age, and seniors.

It would appear then that neither geographic distribution nor sorting under an antagonistic atmosphere produced a noticeable effect. This furnishes us with some evidence on the problem of the present research, since prejudiced subjects seemed to sort about the same as the rest. This is especially significant when it is considered that racial attitudes are perhaps the most emotionally intense of all attitudes, and might be most responsive to sectional bias and emotional heat.

For purposes of tabulation and comparison, the group of Negro subjects was called Group III. The results were recorded and curves drawn, in the same manner as previously described for the other groups. The scale values and Q values were computed for all the statements from these tables and curves, and were properly recorded along with those of the other groups. The Pearson coefficient of correlation between the series of scale values of Group I and those of Group III was then computed, for purposes of comparison.

After a critical examination of the curves and data of Group I, according to Thurstone's method described in "Attitudes Can Be Measured," 32 statements of opinion were chosen as the landmarks in the final scale. It would be possible, in a similar manner, to select a series of evenly spaced statements on the basis of the data for each of the other two groups. It is quite certain that the same 32 statements would not be selected in each case, because of the differences in scale values, although many of the statements would probably be common to all three scales. We would then have three separate scales for measuring attitude toward the Negro, constructed independently on three different groups of subjects. Since the construction of these three scales would have no bearing on the present problem, only one such scale is presented in this paper, although the writer selected the other two series of statements for his own interest. A better comparison than the construction of three scales is obtained by listing the three sets of scale values and Q values for the 32 statements of the first scale, and also by plotting the three sets on three separate base lines. These methods of comparison were used and the differences will be commented upon later.

For further comparison, the actual differences between the scale values of Group I and Group II were calculated. A frequency distribution of these differences, Group I minus Group II, was made, and the mean and standard deviation of this distribution computed. The relationship between these differences and the scale values of

Group I were also studied, as being of possible value to our study. Further proof as to the validity of the scale was obtained by computing the probable error of a scale value in general. This was obtained on the basis of the average Q value, as follows:

$$\begin{aligned}
 \text{Mean } Q &= .97 \\
 Q &= .6745\sigma \\
 \sigma &= 1.4825Q \\
 &= 1.4825(.97) \\
 &= 1.438 \\
 P.E. (\text{median}) &= \frac{.84535\sigma_x}{\sqrt{N}} \\
 &= \frac{.84535(1.438)}{\sqrt{114}} \\
 &= .114
 \end{aligned}$$

This is the probable error of a scale value in general. All of these comparisons and results will be commented on and interpreted in the "Summary and Conclusions."

SUMMARY AND CONCLUSIONS

Following Thurstone's method of attitude scale construction, three sets of scale values for 114 statements have been independently obtained from the sortings of three groups of subjects. Since these groups differ definitely in their opinions as to the social position of the Negro, the crux of our problem lies in the comparison of these sets of scale values. Following are the results of such a comparison:

1. Remarkably close agreement is shown between the curves for Groups I and II, as drawn on the same coordinate axes.

2. The correlation between the scale values of Group I and the scale values of Group II gives a highly linear plot, and a Pearson coefficient of .980. The linearity is evident from the scatter diagram, and it may also be calculated on the basis of the correlation ratios. $\eta_{xy} = .984$; $\eta_{yx} = .985$.

3. The correlation between the scale values of Group I and the scale values of Group III likewise gives a plot which is well within the limits of linearity. $r = .935$; $\eta_{xy} = .947$; $\eta_{yx} = .952$.

4. A comparison of the scale values which the various groups give

to the 32 statements of the final scale throws some light on our problem. Although these statements were chosen because of their scale values in Group I, it is significant that Group II places them in the same rank order in which they were placed by Group I, with one exception. Statement 95 is the only one of the 32 statements which has a different relative position in Group II from what it has in Group I. This means that our scale, which was constructed on the basis of the sortings of Group I, can measure the attitudes of Group II as well as it can the attitudes of Group I. This is very definite evidence in favor of the objective validity of the scale.

5. The very low value obtained for the probable error of a scale value in general gives us increased confidence in the calculated scale values.

6. The frequency distribution of differences between the scale values of Group I and the scale values of Group II gives a mean of .27 and a standard deviation of .55. This constant error reveals a slight tendency for Group I, which is prejudiced against the Negro, to judge a given statement to be more favorable to the Negro than Group II judged it. When we tabulate the differences between the scale values of Groups I and II for only the 32 statements in the completed scale, we obtain a mean of only .17. This is only one and a half times the probable error of a scale value, and could very possibly be due to the fluctuations of chance. At any rate, the fact that the correlation between the two groups is so high, together with the other evidence presented, suggests that this constant error is not significant.

7. The scatter diagram of the relation between the differences (Group I — Group II) and the scale values of Group I show no tendency for the differences to vary with the scale value.

8. It will be recalled that one of the Negro colleges became emotionally wrought up concerning the sorting task, but that their sorting was about the same as that of the other subjects. This furnishes further evidence in favor of the fact that emotional prejudice does not influence the differentiating judgment of the sorter.

On the basis of these points of comparison, we conclude that the scale which we have constructed for measuring attitude toward the social position of the Negro is not influenced in its measuring function by the attitudes of the subjects used in the construction. We

have definitely shown this conclusion to be true for those districts where the social position of the Negro is an issue. It is possible that this would not hold for countries where the social position of the Negro is not an issue. We conclude, in general, that any scale of attitudes, carefully constructed in accordance with Thurstone's method, will not be affected in its measuring function by the position which the sorters occupy on the scale.

It is to be noted that we have been concerned here only with the manner in which the position of the sorters on an attitude scale affects the validity of that scale. We must not confuse this with another factor, which may possibly affect the validity of an attitude scale. This factor is the difference in the standards of evaluation of the statements used by different sorters. To sit down and eat with a Negro might indicate the height of social recognition to one sorter, and indicate a less extreme attitude to another sorter who lives in a community where eating with Negroes is customary. This may be true even though the two sorters occupy approximately the same position on the attitude scale. It might therefore be possible, for a given series of statements, to obtain two quite different sets of scale values, based upon the sortings of two groups who differ in their standards of evaluation of the statements. Although this problem is beyond the scope of the present research, the writer hopes soon to study the influence of this factor upon the validity of an attitude scale.

REFERENCES

1. DROBA, D. Measurement of attitude toward Militarism-Pacifism. (Unpublished.)
2. SMITH, H. Measurement of attitude toward prohibition. (Unpublished.)
3. THURSTONE, L. L. Attitudes can be measured. *Amer. J. Sociol.*, 1928, **33**, 529-554.
4. ———. Law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273-286.
5. ———. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368-389.
6. ———. The measurement of opinion. *J. Abn. & Soc. Psychol.*, 1928, **22**, 415-430.
7. ———. The Phi Gamma hypothesis. *J. Exper. Psychol.*, 1928, **11**, 293-305.
8. ———. Theory of attitude measurement. *Psychol. Rev.*, 1929, **36**, 222-241.
9. ———. Three psychophysical laws. *Psychol. Rev.*, 1927, **34**, 424-432.

10. ———. Unit of measurement in educational scales. *J. Educ. Psychol.*, 1927, 18, 505-524.
11. THURSTONE, L. L. & CHAVE, E. J. Measurement of attitude toward the church. (Unpublished.)

University of Florida
Gainesville, Florida

L'INFLUENCE DE L'OPINION INDIVIDUELLE SUR LA CONSTRUCTION D'UNE ÉCHELLE DES ATTITUDES

(Résumé)

Après un choix très soigneux de 114 expressions d'opinion à l'égard de la position sociale du nègre, on a obtenu indépendamment trois groupes de valeurs d'échelle pour ces expressions de la classification de trois groupes de sujets. Le Groupe I a été un groupe de trois sujets de race blanche du Midi des Etats-Unis, ayant un préjugé contre le nègre. Le Groupe II a été un groupe de sujets de race blanche du Nord des Etats-Unis, ayant un préjugé en faveur du nègre. Le Groupe III a été un groupe de sujets nègres. On a computed chaque groupe de valeurs d'échelle selon la méthode de Thurstone, et l'on a comparé soigneusement les trois groupes.

Dans la comparaison les valeurs d'échelle des Groupes I et II, $r=0,980$, $\eta_{xy}=0,984$, et $\eta_{yx}=0,985$. Dans la comparaison des valeurs d'échelle des Groupes I et III, $r=0,935$, $\eta_{xy}=0,947$, et $\eta_{yx}=0,952$. Ces résultats avec d'autres points de comparaison, le rendent évident que l'échelle que l'on a construite pour mesurer l'attitude à la position sociale du nègre n'est pas influencée dans sa fonction de mesure par les attitudes des sujets employées dans la construction. On conclut en général qu'aucune échelle des attitudes, construite soigneusement selon la méthode de Thurstone, ne sera influencée dans sa fonction de mesure par la position qu'occupent les sujets sur l'échelle.

HINCKLEY

DIE EINWIRKUNG DER PERSÖNLICHEN MEINUNG AUF DIE KONSTRUCTION EINES MASSSTABES ZUR MESSUNG DER EINSTELLUNG

(Referat)

Nach sehr sorgfältiger Auswahl von 114 Meinungsbehauptungen (state-ments of opinion) über die soziale Stellung der Neger wurden drei Gruppen von Rangwerten (scale values) für diese Behauptungen aus den Rangierungen (sortings) von drei Gruppen von Versuchspersonen separat erhalten. Gruppe I bestand aus weissen Versuchspersonen aus dem Süden die eine vorgefasste Meinung gegen die Neger hatten. Die Gruppe II bestand aus weissen Versuchspersonen aus dem Norden die eine vorgefasste Meinung zu Gunsten der Neger hatten. Die Gruppe III bestand aus Negern. Die Werte in jeder Sammlung von Rangwerten wurden nach

der Methode von Thurstone berechnet, und die drei Wertgruppen wurden sorgfältig verglichen.

In der Vergleichung der Rangwerten der Gruppen I und II betrug die Korrelation (r) .980; η_{xy} betrug .984, und η_{yx} betrug .985. In der Vergleichung der Rangwerten der Gruppen I und III betrug die Korrelation (r) .935; η_{xy} betrug .947, und η_{yx} betrug .952. Diese Befunde, im Zusammenhang mit anderen Vergleichen, beweisen, dass der Massstab (scale) wir zur Messung der Einstellung der sozialen Lage des Negers gegenüber konstruiert haben, durch die Einstellungen der in der Konstruierung der Massstabes begrauchten Versuchspersonen in seiner Funktion der Messung nicht beeinflusst wird. Wir ziehen hieraus den allgemeinen Schluss, dass irgend ein Massstab zur Messung von Einstellungen, wenn er sorgfältig nach der Methode von Thurstone konstruiert wird, durch die Lagen, die von den verschiedenen Auslesern (sorters) auf dem Massstab eingenommen werden in seiner Funktion der Messung nicht beeinflusst werden wird.

HINCKLEY

A QUANTITATIVE ANALYSIS OF MARRIAGE SELECTION IN A SMALL GROUP*¹

From the Psychological Laboratories of Columbia University

BELLE SCHILLER

Of recent years increasing attention has been given to the factors entering into the establishment and maintenance of the marriage relationship. Since the subject is one of universal concern, proponents from very different fields have taken up the discussion. They may, however, be divided into two main classes, the preceptive and the experimental. The former is concerned with defending or attacking existing mores, and with the factors bearing upon the conduct and continuance rather than the establishment of the marriage. Since the evolution of social institutions is too slow a process for any immediate noteworthy change to be expected, regardless of the extent of dissatisfaction or discussion, we may, with more profit, perhaps, turn to those studies which are concerned with the analysis of the factors entering into the establishment of the marriage, for not until we have some knowledge of the forces at work in the selection of mates should we attempt any generalizations concerning the relationship itself. To obtain such knowledge will be a difficult enough task. The occasional instances in which the attraction between two individuals breaks through differences in age, race, color, education, religion, social or cultural status—in short, the habits, the attitudes, the ties and affections of a lifetime—would be enough to give us pause, were it not that in the greater number of marriages selection is less spectacular and more open to objective measurement.

Such measurement was first attempted by Francis Galton, who in 1869, after a study of "nearly 1000" eminent men involving some three hundred families, concluded that his data established the existence of a tendency of "like to like" among intellectual men and women" (5, p. 325). In 1889 he attacked the problem more

*Accepted for publication by A. T. Poffenberger of the Editorial Board and received in the Editorial Office, June 3, 1931.

¹The writer is deeply indebted to Professor H. E. Garrett, Columbia University, for the interest, constant guidance, and helpful criticism which made this paper possible.

directly in his *Natural Inheritance* and decided that there was little evidence of marriage selection in stature, "temper," eye color and artistic tastes in his two hundred cases (6, p. 85). His data were later used by Pearson and his co-workers who made a more statistically adequate study of the subject under the term of assortative mating which they defined as "the attraction of certain classes of males to certain classes of females" (24, p. 481).

This study has not been limited to man. Pearl, working with paramecia, found that the "true value of the correlation coefficient measuring assortative pairing in respect to length lies between .5 and .6" (15, p. 275). D. F. Jones (9) reports several such investigations. Jennings, for instance, obtained an r of .25 for length in a race of paramecia descending from a single individual, and one of .38 in a mixed race. Crozier found a coefficient of .6 for the mollusk *Chromodoris zebra* (a hermaphroditic nudibranch). Crozier and Snyder report for crustacea a coefficient of $.91 \pm .01$ for *Gammarus locusta* and one of $.69 \pm .04$ for *Dikero gammarus fasceatus*. Among insects similar evidence has been obtained for *Leptinotarsa* (potato-beetle) (9, pp. 119-131). These data, scanty as they are, are nevertheless important on two accounts; they agree on the positive character of the coefficient, and they indicate the phylogenetic significance of the problem.

In man the investigation of assortative mating is of interest in four fields: genetics, eugenics, sociology, and psychology. Parent-child correlations, for example, must be interpreted in the light of the parental coefficients. In fact, techniques in child psychology remain inadequate as long as data regarding the parents, such data as are necessarily involved in a study of assortative mating, are neglected. H. E. Jones (11), in his recent survey of the literature, considers the problem one of five-fold significance, requiring consideration in connection with the "psychological theory of types, the study of the origin and maintenance of social classes, in the statistical analysis of heredity, in psychoanalytical and other studies of social relationships in the home, and in the investigation of marital infection . . . and in evolutionary and eugenical theory." In short, the problem of marriage selection is important, not only because the marriage relationship involves the total behavioral organization of the individual, but because that relationship is a basic and all-pervasive factor in the life of the group.

Survey of Literature. Research in assortative mating has con-

cerned itself with the physical, mental, and temperamental or emotional² characteristics of husband and wife and will be tabulated under these heads.³

These investigations, differing as they do in methods, subjects, and results, are characterized by certain similarities. The coefficients for the mental and physical characteristics, although they differ in size, are uniformly positive, while the temperamental characteristics are, on the whole, negative as regards selection. In so far as method is concerned, a systematic study of the same group has not been made. In addition, although there is some evidence of a limitation of subjects to a more homogeneous and therefore more accurate sampling of the population under consideration, in no case is such limitation rigorous or consistent enough, for to the extent that the groups tested have been heterogeneous with respect to age, education or culture, racial, religious, social, or economic factors, the correlations based on such groups have been distorted. Finally, except for some first-cousin correlations obtained by Pearson and his workers (17, p. 410), the use of some sort of control technique has been avoided.

It is the aim of the present study to determine in a strictly homogeneous group the extent to which assortative mating exists with respect to as many factors as could be studied in the amount of time the group was willing to give, and to compare the extent of assortative mating as measured by correlation coefficients thus obtained with coefficients obtained by a random sampling of the same group. In insisting on strict homogeneity three results have been achieved: the disturbing effect of the presence of extremes on the correlation coefficients have been avoided, the group used is very representative of its particular segment of the whole population, and the correlations obtained for the control group are more significant.

Description of Subjects. Of the 51 couples tested,⁴ 46 were finally selected as being homogeneous enough in age, education, occupation, socio-economic status, length of marriage, and religion to be included in the group. The average duration of marriage was 2.534 years, with a standard deviation of 2.32 and a range of .08 to 10.58 years. Forty-four of the marriages fell between one month and

²Studies have also been made of certain actuarial and pathological aspects of assortative mating, but, since these are not directly relevant to our problem, they will not be discussed here.

³Based on Elderton's tabulation of the then existent data, *Draper's Co. Res. Mem.*, 1908, 3.

⁴During June and the first two weeks in July, 1930.

TABLE 1
SURVEY OF PREVIOUS STUDIES OF MARRIAGE SELECTION

Characteristic	Investigator	No. of couples	Correlation	Remarks
<i>Physical</i>				
Stature and eye color	Galton (6, p. 85)	205	No <i>r</i> computed	Data analyzed in terms of average and variability. Little evidence of marriage selection.
Stature	Pearson (16, p. 270)	205	.0931±.0473	From Galton's data, upper middle classes.
Eye color	Pearson, Lee (23, p. 113)	205	.1002±.0378	From Galton's data, upper middle classes.
Eye color	Pearson (20, p. 475)	774	.2600±.0300	Mean square contingency.
Stature	Pearson, Lee (22, p. 373)	1,079	.2804±.0189	Subjects "very largely from the professional classes" (18, p. 28).
Span		1,005	.1989±.1204	
Forearm		1,000	.1977±.0205	
Husband's stature, wife's span		1,055	.1820±.0201	
Husband's stature, wife's forearm		1,050	.1403±.0204	
Husband's span, wife's stature		1,053	.2023±.0199	
Husband's span, wife's forearm		1,053	.1533±.0203	
Husband's forearm, wife's stature		1,051	.1784±.0201	
Husband's forearm, wife's span		1,051	.1545±.0203	

TABLE 1 (*Continued*)

Characteristic	Investigator	No. of couples	Correlation	Remarks
Stature	Davenport (2, p. 329)	137	No r computed	Method based on probability. Persons of similar stature tend to marry each other. Extremes more particular in this respect.
Body build	Davenport (2a, p. 46)	531		"Some degree of assortative mating....of similars."
Age	Lutz (14)	2,500	.7640	Records of Chicago marriage license office.
Age	Pearson, Heron (21, pp. 224-225) *	5,317,520	.9250	British census for 1901.

*Coefficients of contingency for this data range from .84 to .91 depending on fineness of grouping.

TABLE 1 (Continued)

Characteristic	Investigator	No. of couples	Correlation
<i>Mental test ability</i>			
"Ability"			
Association	Galton (5, p. 315)	300 families	Tendency of "like to like" established.
	Fürst (4, p. 444), Jung (12, p. 166)	9 families†	Average difference of 4.7.‡
Army Alpha	Jones (10)	105	Reliability .977§
			<i>r</i>
			Corrected <i>r</i>
Opposites, N. I. T.	Willoughby (30, pp. 260-269)	90 (?)	H W
Number Series, Army Alpha			.92 .90
Arithmetic Reasoning, N. I. T.			.75 .94
Symbol Series Army Beta			.68 .98
Vocabulary Recognition, Stan- ford Achievement			.92 .91
Form Combination, Army Beta			.84 .88
Analogies, N. I. T.			.73 .79
Symbol Digit, Army Beta			.88 .86
Science Nature Information,			.85 .88
Stanford Achievement			.56±.05
History Literature Informa- tion, Stanford Achievement			.43±.05
Comparison, N. I. T.			.95 .92
			.55±.06
			.41±.06
			.46±.05
			.58±.06
			.45±.06

†Comprising 37 individuals with an age range of 9 to 74. The number of married pairs is not cited. It is doubtful whether there were more than ten.

‡Jung gives the maximum possible difference as 13.3. His obtained difference therefore amounts to 35% or 65% agreement. Fürst, in her more detailed treatment of the same data, obtains an A.D. of 3.2. She characterizes the 4.7 as "a fairly slight average agreement which however varies greatly, i.e., there are cases of higher agreement and others of the greatest divergence."

§Calculated for 102 tests taken at random.

||Number of couples not stated. May be approximated from the fact that there were 100 mothers and 90 fathers involved.

TABLE 1 (*Continued*)

Characteristic	Investigator	No. of couples	Correlation
<i>Temperamental</i>			
Artistic taste and temper	Galton (6, p. 85)	205	Little evidence of marriage selection.
Hyperkinetic, hypokinetic, normal temperament.	Davenport (1, p. 108)	100 family histories	Marriage selection against similar temperaments.
		Mean square contingency	Fourfold correlation
Temperament, merry or not	Schuster, Elderton (26, pp. 467-468)	.19	-.05
Tendency towards drink		.27	.36
Political opinions		.29	.36
Truthfulness		.22	-.04
Religious feeling		.53	.73
Manner of speech		.30	.11
Tone of voice		.22	.15
Nature with regard to money expenditure		.34	.17
Full of movement or sedate and quiet		-.23	-.23
Occupies his leisure moments or inclined to make himself comfortable		-.03	-.03
Neglects duties for other occupations or not			.20
Resolute or undecided			-.03
Critical or idealizing			-.29
Anxious or sanguine			-.28
Decided or cautious in the expression of opinion			-.02
Caring much for good eating or drinking or not			.42
Reading much or not			.11
Emotional or not			-.17

4.42 years. The average age of the husbands was 27.672 years and that of the wives, 24.804. As far as formal education was concerned, all except one of the men and three of the women were high-school graduates; the degrees held by the rest of the former included B.A., B.S., M.A., Ph.D., Ph.G., M.D., D.D.S., LL.B., and C.E., and by the latter B.A., B.S., M.A., and LL.B. Some were normal-school graduates and several were candidates for the Ph.D.

The occupations showed greater variety. Six men were engaged in business, eight in teaching (the positions ranged from elementary school to college), and the rest were professionals, physicians, dentists, engineers, lawyers (eleven), research workers, social workers, students, etc. Half of the wives were teachers (again elementary school to college), ten were housewives, four were professionals, (biochemist, lawyers, etc.), and the rest were clerical workers or indicated no vocation.

In addition, the group was composed in the main of first generation Russian, Austrian, and to a smaller extent Polish and German Jewish immigrants. The social status of the group is indicated by its educational and occupational range. The last-named factor and the average age are standards by which the economic level of the group may be gauged, as it was deemed inexpedient to ask for information concerning salary. In most cases both husband and wife were wage-earners. In some cases this was due to necessity, in others, to the wife's refusal to be burdened with the care of a home and children. Several of the wives, however, intimated plans to "settle down" as soon as the husband had obtained an adequate livelihood. Seven children were recorded (three by one couple and two by another, and one each by two couples), and three wives mentioned pregnancy.

Attempts were made to form a control group by pairing at random the test results for an equal number of unmarried men and women. It was impossible, however, to assemble a comparable group, since the unmarried subjects were considerably younger, with education not so far advanced and the economical and occupational status less firmly established. A system⁵ of random matings from among the

⁵The tests having already been numbered (husband and wife were given the same number, e.g., H1, W1 for the first pair, etc.) in order to overcome the objections of potential subjects who feared that the tests would be too difficult or that intimate personal questions would be asked, it was merely necessary to shuffle the tests so that no individual was paired with his own mate. This was done entirely at random.

married group was therefore resorted to, and the matings thus formed used throughout as the control group.

Description of Tests. The selection of the tests was conditioned by two factors, the educational status of the group and the amount of time it was willing to give. Because of the first factor it was necessary to give, in so far as the mental tests were concerned, such tests as the subjects were unfamiliar with despite college courses in psychology, as well as difficult enough to be stimulating. So-called general intelligence tests were ruled out on the assumption that it would be more advisable to test specific and recognizable mental abilities. Since preliminary questioning revealed an unwillingness on the part of the subjects to give more than an hour and a quarter, it was necessary to shorten the original tests in order to obtain as wide a cross-section of personality make-up as possible and thus provide more bases for comparison.

The tests were taken by the subjects in their homes, usually in the evening. Often a couple had invited married friends, so that several couples were tested at once. This circumstance, as well as the fact that names were not required, helped to eliminate tension.

Before they were tested the subjects filled out the following questionnaire:⁶

Please fill in the following items as accurately as you can. All individual information will be held confidential. This investigation is concerned only with group results.

Identification (H or W Number....)....Date of Marriage....

Date of birth.....Order of birth.....

Number of older brothers.....Number of older sisters.....

Number of younger brothers....Number of younger sisters....

Length of acquaintance before marriage.....

Length of acquaintance before engagement.....

Height Weight Hair color

Indicate eye color—Gray.... Blue.... Green.... Brown....

Father's height.... Father's weight.... Father's hair color....

Father's eye color—Gray.... Blue.... Green.... Brown....

Mother's height.... Mother's weight.... Mother's hair color....

Mother's eye color—Gray.... Blue.... Green.... Brown....

Schooling: Indicate graduation or length of attendance

Elementary school High school.....

⁶Questions were also asked concerning favorite subject at school, major and minor at college, and recreational interests, but this information was not utilized.

Normal school	College
Further details concerning schooling	
Occupation	
Occupation before marriage	

A description of the tests follows in the order in which they were given:

1. An association test composed of 50 alternate words from the Kent-Rosanoff list (13a).
2. A 10-minute arithmetic reasoning test composed of 20 items from M. M. R. Schneck's Arithmetic Reasoning Test (25).
3. A 10-minute multiple-choice vocabulary test consisting of 90 items from Schneck's Vocabulary Test and an additional 10 extremely difficult words selected from the December 1927 examination for elementary school principalship in New York City.⁷
4. A 10-minute, multiple-choice information test containing 40 items of literary and scientific interest from this last-named source.⁷
5. Whitman's Shortened Extroversion-Introversion Scale (29). This consists of 10 items from the Colgate Personal Inventory Schedule C2 which correlate $.81 \pm .025$ with the whole test.
6. 28 items from the Woodworth-House Mental Hygiene Inventory (8).⁸ These were selected on the basis of the following criteria: avoidance of physical and sexual topics, omission of questions relating to childhood experiences, and the inclusion of as varied an assortment of items as possible.
7. 30 items from a public-opinion questionnaire now being developed at Columbia University, modified to permit Yes and No answers. They embody matters of economic, racial, religious, industrial and social interest.

Results

Physical. As the questionnaire indicates, the subjects were asked for the height, weight, hair and eye color of their parents as well as their own. Correlations were then computed for husband-wife's father and wife-husband's mother, in order to test, in these physical aspects at least, the Freudian assumption concerning marriage selection.⁹ This assumption has not been confined to the psychoanalysts, for Pearson, in 1900, ventured the statement, "A

⁷New York Sun, Home Edition, Dec. 31, 1927; Jan. 7, 9, 10, 1928.

⁸A modification of the Woodworth Personal Data Sheet allowing for an extreme and a moderate form.

⁹See especially (3, p. 27).

MARRIAGE SELECTION IN A SMALL GROUP

307

Characteristic	Range	Average	S.D.	V*	Diff.		Random
					S.D.	diff.	
Age							
Husbands	22-39	27.672	3.245	1.173			
Wives	21-35	24.804	2.716	1.055	8.11	.689±.052	.058±.098
Height							
Husbands	62-72	68.239	2.722	3.989			
Wives	59-67	62.782	2.715	4.325	12.01	.355±.086	.097±.098
Husbands	62-72	68.239	2.722	3.989			
Wife's father	60-72	66.152	2.585	3.908	5.02	.434±.080	-.817±.033
Wife	59-67	62.782	2.715	4.325			
Husband's mother	56-69	63.152	2.680	4.243	.85	.370±.085	.164±.096
Weight							
Husband†	123-230	157.326	20.050	12.744			
Wives	93-150	122.934	13.115	10.668	11.62	.321±.088	-.119±.097
Husband†	135-220	155.222	17.578	11.303			
Wives	83-140	119.222	13.454	11.285	11.04	.428±.106	.011±.130
Husband§	125-230	161.444	20.445	12.664			
Wives	93-150	124.629	13.601	10.913	9.37	.334±.115	-.732±.047
Husband	123-230	157.659	19.750	12.527			
Wife's father	128-220	159.772	16.703	10.454	.61	.214±.097	-.024±.102
Wife¶	93-150	123.818	12.329	9.957			
Husband's mother	110-200	150.386	20.664	13.741	8.10	.205±.097	-.631±.061
Order of birth							
Husbands	1-8	2.587	1.715	6.629			
Wives	1-8	2.239	1.797	8.026	.13	.067±.098	.138±.010
Number of siblings							
Husbands	0-8	3.521	1.965	5.581			
Wives	0-8	2.978	1.715	5.759	.69	.120±.097	.114±.097

*Coefficient of variation, cf. Garrett (7, pp. 40-45).

†46 couples now. §Same 27 couples now.

||44 cases.

¶27 couples who gave weight at time of engagement

||43 cases.

TABLE 3
FURTHER RESULTS FOR PHYSICAL TESTS

Characteristic	Married		Random	
	Percentage agreement*	C†	Percentage agreement	C
Hair color				
Husband-wife	54.3	.206	58.7	.147
Husband-wife's mother	60.0	.353	50.0‡	.249
Wife-husband's mother	58.1§	.235	60.5§	.207
Eye color				
Husband-wife	45.7	.416	32.6	.401
Husband-wife's mother	45.7	.446	43.5	.339
Wife-husband's mother	52.2	.481	34.8	.324

*Percentage of cases in which hair and eyes were of the same color.

†Coefficient of contingency calculated for a threefold classification for the hair—fair, brown, black; and for a fourfold classification for the eyes—gray, blue, green, brown. The maximum coefficient for a threefold classification is .316; for a fourfold, .966.

‡44 cases.

§43 cases.

correlation, if there be substantial selective mating, may even exist between a man and his mother-in-law" (19, p. 46).

On the basis of Tables 2 and 3 the following statements may be made:

1. The age coefficient of $.689 \pm .052$ is the highest one obtained and points to a definite selective tendency in this respect. It is especially valuable inasmuch as the coefficient for the random group is $.058 \pm .098$.

2. The height and weight coefficients of $.355 \pm .086$ and $.321 \pm .088$ are approximately equal, and both, because of their absolute size and their random r 's indicate a smaller, but probably trustworthy, assortative tendency. Since the coefficient is $.428 \pm .106$ for the 27 couples who gave their weight at the time of engagement, it appears that this tendency is stronger at the time when selection takes place. Since the amount of increase in weight after marriage undoubtedly differs for different individuals, the effect on the correlation coefficient is difficult to estimate. It is rather interesting that for r for the 27 couples "now," $.334 \pm .115$, is very similar to the r of $.321 \pm .088$ for all 46 couples; a similarity due undoubtedly to the homogeneity of the group.

3. The wives show a stronger tendency to select husbands as tall as their fathers than do the husbands to select wives as tall as their mothers. The in-law correlations of $.314 \pm .097$ and $.205 \pm .097$ for weight are indeterminably affected by the change in weight with age. In view of the negative character of the coefficients for the control in-law correlations, however, it is possible that there is an assortative tendency in this respect.

4. As regards order of birth there appears to be no assortative tendency, since the r of $.138 \pm .010$, though low, is distinctly higher than that of $.067 \pm .098$ for the married. Nevertheless, seven couples were both first-born and four couples both second-born.

5. The married and the control group give approximately equal coefficients for the number of siblings, $.120 \pm .098$ and $.114 \pm .097$. There seems to be no selection in this respect.

6. The results for hair and eye color are not so clear cut. The coefficients of contingency are higher for eye color than for hair color, but the coefficients for the random group are almost as high. It seems certain that because the group was an all-Jewish one the range was too narrow and the distribution too centralized for the random matings to produce very much lower coefficients. For this reason, too, a selective tendency would have but a limited chance to operate. Nevertheless, small though the difference is, the coefficients are in all cases higher for the married group.

7. As regards variability, husbands and wives are about equal. Neither is uniformly more variable.

8. The differences in age, height and weight are definitely reliable for husbands and wives, while those for order of birth, number of siblings are as definitely unreliable. In height, the difference between husband-wife's father, and in weight that between wife-husband's mother are reliable.

9. These results and those of previous workers indicate an assortative tendency in physical characteristics, although there is no definite agreement as to the strength of this tendency.

Mental. In Table 4 the r corrected for attenuation is given first, the raw coefficient follows directly beneath it in parenthesis.

Because of the unreliability of the information test it will be discussed separately later. Of the arithmetic reasoning and vocabulary tests the following statements may be made:

1. The husbands are in both cases superior to the wives, and the difference is reliable for the arithmetic reasoning, and fairly high

TABLE 4
RESULTS FOR ARITHMETIC REASONING, VOCABULARY, AND INFORMATION TESTS

Test	Range	Average	S.D.	V	$\frac{\text{Diff.}}{\text{S.D.}}$	r	Random r	Reliability coefficients*
Arithmetic Reasoning Husbands Wives	4-20	11.348	3.856	33.98	5.73	.344 \pm .088 (.293 \pm .090)	-.138 \pm .097 (-.118 \pm .097)	.862 .844
	3-19	7.761	3.245	41.85				
Vocabulary Husbands Wives	5-83	42.521	19.82	46.589	2.23	.117 \pm .098 (.109 \pm .097)	.086 \pm .098 (-.080 \pm .098)	.963 .896
	12-61	34.891	14.56	41.729				
Information Husbands Wives	8-35	18.562	5.684	30.622	3.85	(.414 \pm .082)	(-.077 \pm .099)	.663 .669
	2-26	15.283	4.918	32.186				

*Computed by the split-half method (alternate items). The coefficients for arithmetic reasoning and vocabulary tests compare favorably with the original. The information test proved unreliable, probably because the items had not been equated for difficulty or arranged in order of difficulty in the construction of the test, proceedings which are essential for the use of the split-half method. Because of the unreliability, no correction has been made for attenuation, in this test. The figures given represent Brown-Spearman correction, the original self-correlations were, respectively, .757, .730; .927, .811; .496, .470.

though not completely reliable for the vocabulary test. This superiority may be due to the fact that the husbands had more schooling.

2. The correlation is higher for the arithmetic reasoning than for the vocabulary. The corrected r of .344 is fairly substantial, and, in view of the negative character of the coefficient for the random group, is especially significant.

3. The correlation in vocabulary is negligible. For this result the size of the *S.D.* is to some extent responsible. The test is an extremely difficult one.

4. The corrected coefficient for arithmetic reasoning is almost identical with Willoughby's raw coefficient of .34, but he reports a much higher r for his vocabulary test than the one obtained here.

5. As regards the information test, the husbands are again superior, the difference is reliable, and the raw coefficient the highest obtained for the three tests, .414. Of Willoughby's two analogous tests, the Science Nature Information gave a correlation of .46¹⁰ and History Literature Information .55.¹⁰

Association. The reactions were classified according to Well's (28)¹¹ fivefold classification. The number of responses of each type was found for each husband and wife and correlation coefficients calculated for each of the five classes of association.

The following conclusions may be drawn concerning the association reactions:

1. The averages, *S.D.*'s and coefficients of variability are almost equal for husbands and wives.

2. The contrast, speech-habit, and egocentric classifications give the largest coefficients. Husband and wife tend to give the same number of responses which reflect attitudes and language habits, for the coefficients for the contrast and speech-habit are fairly high, $.468 \pm .079$ and $.464 \pm .080$, not only in absolute size, but in relation to the *P.E.* and to the random coefficient. If we consider the egocentric responses as responses "that show a special participation of subject's personal experiences and interests" (31, p. 125) then it may

¹⁰Corrected for attenuation.

¹¹Predicate or egocentric ("special subjectivity of attitude"—proper nouns, pronouns, failure or repetition of stimulus words); supraordinate (individual-genus only); contrast; miscellaneous or inner (causalities, coordinates, subordinates, identities, and doubtful;—"the predominance of this heterogeneous group indicates a 'concrete' reaction tendency as distinguished from an egocentric one"); and speech-habit associations.

TABLE 5
RESULTS FOR KENT-ROSANOFF ASSOCIATION TEST

Class of association	Range	Average	S.D.	<i>V</i>	$\frac{\text{Diff.}}{\text{S.D. diff.}}$	<i>r</i>	Random <i>r</i>
Egocentric Husband Wife	2-37 5-42	20.200 21.886	8.075 8.856	39.98 40.46	1.21	.378±.087	— .390±.086
Supraordinate Husband Wife	0-6 0-7	1.318 1.023	1.755 1.453	133.16 142.03	.31	.222±.097	.158±.099
Contrast Husband Wife	0-17 0-18	7.091 5.955	5.648 4.967	79.650 83.409	1.40	.468±.079	.041±.102
Inner Husband Wife	3-31 7-27	16.705 17.316	5.801 5.341	34.762 31.368	.68	.041±.102	— .461±.080
Speech-Habit Husband Wife	0-15 0-13	4.500 4.000	3.428 2.804	76.178 70.100	1.04	.464±.080	— .121±.100

TABLE 6

Test	Range	Average	S.D.	<i>V</i>	<i>Diff.</i>		<i>r</i>	Random <i>r</i>
					<i>S.D.</i>	<i>diff.</i>		
Extroversion-Introversion								
Husband	1.50-4.1	3.125	.484	15.480	3.28	.005±.098		.223±.093
Wife	2.44-4.3	3.429	.406	11.840				
Mental Hygiene Inventory								
Husband	0-26	11.196	5.491	49.04	2.45	.021±.098		.059±.098
Wife	2-26	14.043	5.802	41.315				

be concluded that husband and wife tend to give the same number of responses based on personal experience and interest, since the coefficient for the number of egocentric responses, though smaller, $.378 \pm .087$, is still reliable in terms of the *P.E.* and random *r*.

3. The supraordinate and the inner types of association give lower correlations, $.222 \pm .097$ and $.041 \pm .102$. Since these responses are more generalized, objective, and abstract¹² than the others, and therefore less apt to be evoked by experiences capable of being shared or communicated, a low degree of correlation (as in the case of the supraordinates) or even absence of correlation (inner responses) is perhaps to be expected.

4. With the exception of the supraordinates, the results for the random group are very small or decidedly negative.

5. These findings agree with those of Jung and Fürst as to the existence of a general tendency towards similarity.

It may be concluded, as in the case of the physical characteristics, that a selective tendency exists for such mental traits as have been measured (with the exception of the result obtained here for vocabulary), although its strength is not agreed upon.

"*Temperamental.*" The responses in the Introversion-Extroversion Scale were rated from 1 for most extroverted to 5 for the most introverted, and the scores for the 10 items averaged for each husband and wife. These average scores were then correlated. The number of "yes" or "wrong" responses on the Mental Hygiene Inventory was found for each husband and wife and a coefficient of correlation computed. As is known, the Woodworth Personal Data Sheet, on which it is very closeley modeled, is a measure of emotional instability.

¹²Cf. Well's definitions.

For the traits measured by these two tests we may conclude:

1. The averages differ reliably for the extroversion-introversion, wives tending to be more introverted and to give an appreciably larger number of "wrong" responses in the Mental Hygiene Inventory, though the difference here is not reliable.
2. The husbands are in both cases more variable.
3. The coefficients for these tests indicate a total absence of correlation.

Opinions on Current Topics. The Yes and No answers to the questions on current topics were thrown into the fourfold tables below (Table 7) and both Shepard's *U*, an approximation of the tetrachoric correlation coefficient (27, p. 171),¹³ and the percentage of items to which husband and wife gave the same response calculated.

The strong tendency of the husbands and wives to answer certain questions in the same way is responsible for the similarity of the results for the married and control groups. For instance, almost every subject answered "No" to the question, "Do we need universal compulsory military training?" and "Yes" to "Do you believe employees should participate in the management of industry?"

TABLE 7
OPINIONS ON CURRENT TOPICS

	Married Wives				Random Wives		
	Yes	No	Total		Yes	No	Total
Husbands				Husbands			
Yes	584	152	736	Yes	544	192	736
No	183	295	478	No	223	255	478
Total	767	447	1214	Total	767	447	1214
<i>U</i>			.649	<i>U</i>			.476
Percentage agreement			72.41	Percentage agreement			55.82

For this tendency of similarity of reaction to topics of industrial, political, social, etc., interest, several factors are responsible. To the homogeneity of the group in economic, religious, occupation, and general social and cultural levels must be added the fact that the subjects are all still young and not apt to be conservative or satis-

¹³Reliability not found as split-half method is not applicable and retesting was not possible.

fied with existing orders. The home environment before marriage was in many cases similar. The greater number of parents were immigrants born in Austria, Russia, Poland, or neighboring countries, of about the same educational and occupational status. With so many factors making for homogeneity and therefore similarity of response, it would indeed have been surprising had the group showed any lesser tendency to agree in its reactions to this questionnaire.

Education, Occupation, and Political Preference. The information concerning education was thrown into a fivefold classification: (1) towards high-school graduation, (2) towards and including college graduation, (3) towards and including Master of Arts degree, (4) towards Ph.D., M.D., LL.B., etc., (5) possessing these degrees; and a coefficient of contingency calculated. The occupations were thrown into a threefold classification, business, teaching, and professional for the husbands, and into a fivefold classification, business, teaching, professional, housewife, and none for the wives. For political preferences, the following classification was used, Republican, Democrat, Socialist, Others (radicals, liberal, etc.), and None.

For occupation and political preference, coefficients of contingency were computed; for the former C was found for occupation previous to marriage as well as occupation now, that is, at the time of testing. Finally, percentages of agreement were calculated for those instances in which education, occupation and political preference were identical for husband and wife.

1. The various factors already discussed which contribute to the homogeneity of the group and stamp it as a closely-knit and well-differentiated stratum of society affect the results here as they have previously. Because of this homogeneity, the range is too small for even random matings to produce low coefficients so that married and control groups give similar results.

TABLE 8
EDUCATION, OCCUPATION, AND POLITICAL PREFERENCE

	Married		Random	
	C	Percentage agreement	C	Percentage agreement
Education	.430	23.91	.571	21.74
Occupation: Previous	.524	19.57	.470	19.57
Now	.523	23.91	.226	13.04
Political preference	.545	52.17	.495	32.61

2. However, except in education, the married group gives higher results than the random. In itself, a coefficient of .420 is large enough to establish the existence of a homogamous tendency, but in comparison with the C of .571 for the random group, the tendency is far from being as strong as it might be and this may explain the anomalous results for the mental tests.

3. The large C for political preference is probably linked up with the similarity of attitude in the current topics questionnaire.

SUMMARY

1. In a highly homogeneous group consisting of 46 married couples, data on certain physical, mental, and temperamental traits were analyzed to discover the extent of assortative mating and were compared with those for a control group formed by a system of random matings.

2. In such physical traits as age, height, weight, hair and eye color, and in husband-wife's father and wife-husband's mother relationship in these traits, assortative mating operates through homogamy. As regards order of birth and number of siblings there appears to be no selection.

3. In such mental traits as association reactions, arithmetic reasoning, vocabulary, information, and opinions, as measured by the tests used, selective mating is again operative through homogamy, while the absence of correlation in vocabulary points to a chance mating, for a negative coefficient would be required to prove heterogamy.

4. In temperamental and emotional traits a chance mating is indicated by the absence of correlation in the tests used.

5. These results agree in trend with previous research in this field, diverse as it has been in subjects and methodology.

6. It should in all cases be remembered that the results in this study, as in others, have been obtained for an already married group. Here the average duration of marriage was 2.534 years, ample time, especially considering the age range—some subjects had married in the late teens or early twenties—for the occurrence of some changes in mental and emotional make-up. Obviously, any assortative tendency which exists is operative at time of selection of mate; the close contact of marriage is apt to affect mental skills as well as attitudes and emotional habit systems. To the extent that an individual changes after marriage, any assortative tendency measured by correlation coefficients obtained at any appreciable interval after marriage is unreliable.

REFERENCES

1. DAVENPORT, C. B. Feebly inhibited: II. Inheritance of temperament. *Publ. Carnegie Instit. Wash.*, 1915, No. 236. Pp. 160.
2. ———. Inheritance of stature. *Genetics*, 1917, **2**, 313-389.
- 2a. ———. Body build and its inheritance. *Proc. Nat. Acad. Sci.*, 1923, **9**, 226-238.
3. FREUD, S. Totem and taboo. New York: Moffat, Yard, 1918. Pp. 265.
4. FÜRST, C. Chap. 11 in *Studies in word association*, by C. G. Jung. New York: Moffat, Yard, 1919. Pp. 575.
5. GALTON, F. Hereditary genius. London & New York: Macmillan, 1869. Pp. 390.
6. ———. Natural inheritance. London & New York: Macmillan, 1889. Pp. 259.
7. GARRETT, H. E. Statistics in psychology and education. New York: Longmans, Green, 1926. Pp. 317.
8. HOUSE, S. D. A mental hygiene inventory. *Arch. Psychol.*, 1927, **14**, No. 88. Pp. 112.
9. JONES, D. F. Selective fertilization. Chicago: Univ. Chicago Press, 1928. Pp. 163.
10. JONES, H. E. A first study of parent-child resemblance. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928, 61-72.
11. ———. Homogamy in mental abilities. *Amer. J. Sociol.*, 1929, **35**, 369-382.
12. JUNG, C. G. Associations d'idées familiales. *Arch. de psychol.*, 1907, **7**, 160-168.
13. KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insan.*, 1910, **67**, 37-96, 317-390.
- 13a. ———. Free association test. Reprint from *Manual of psychiatry*, by A. J. Rosanoff. New York: Wiley, 1920.
14. LUTZ, F. E. Assortative mating in man. *Science*, 1905, **22**, 249-250.
15. PEARL, R. A biometrical study of conjugation in paramecium. *Biometrika*, 1906, **5**, 214-297.
16. PEARSON, K. Mathematical contributions to the theory of evolution: III. Regression, heredity, and panmixia. *Roy. Soc. London, Phil. Trans.*, 1896, **187A**, 253-318.
17. ———. Mathematical contributions to the theory of evolution: on the law of ancestral heredity. *Proc. Roy. Soc. London*, 1897-1898, **62**, 386-412.
18. ———. Data for the problem of evolution in man. *Proc. Roy. Soc. London*, 1899, **66**, 23-32.
19. ———. Mathematical contributions to the theory of evolution: VII. On the correlation of characters not quantitatively measurable. *Roy. Soc. London, Phil. Trans.*, 1900, **195A**, 1-47.
20. ———. Reply to certain criticisms of Mr. G. U. Yule. *Biometrika*, 1906, **5**, 470-476.
21. PEARSON, K., & HERON, D. On theories of association. *Biometrika*, 1913, **9**, 159-315.
22. PEARSON, K., & LEE, A. On the laws of inheritance in man: I. Inheritance of physical characters. *Biometrika*, 1902, **2**, 357-462.

23. ————. Mathematical contributions to the theory of evolution: VIII. On the inheritance of characters not capable of exact quantitative measurement. *Roy. Soc. London, Phil. Trans.*, 1900, **195A**, 79-150.
24. PEARSON, K., & others. Cooperative study of assortative mating in man. *Biometrika*, 1902, **2**, 481-498.
25. SCHNECK, M. M. R. The measurement of verbal and numerical abilities. *Arch. Psychol.*, 1929, **17**, No. 107. Pp. 49.
26. SCHUSTER, E., & ELDETON, E. The inheritance of psychical characters. *Biometrika*, 1906, **5**, 460-469.
27. THORNDIKE, E. L. Mental and social measurements. New York: Teach. Coll., Columbia Univ., 1913. Pp. 271.
28. WELLS, F. L. A preliminary note on the categories of association reactions. *Psychol. Rev.*, 1911, **18**, 229-233.
29. WHITMAN, R. H. A short scale for measuring introversion-extroversion. *J. Appl. Psychol.*, 1929, **13**, 499-504.
30. WILLOUGHBY, R. R. Family similarities in mental-test abilities. *Genet. Psychol. Monog.*, 1927, **2**, 239-277.
31. WOODWORTH, R. S., & POFFENBERGER, A. T. Textbook of experimental psychology. New York, 1920. Pp. 285 (mimeographed).

Brooklyn, New York

L'ANALYSE QUANTITATIVE DU CHOIX D'ÉPOUX DANS UN PETIT GROUPE

(Résumé)

Dans un groupe de quarante-six époux, homogènes à l'égard de l'âge, de l'éducation, de la religion, du métier et de l'état socio-économique, on a analysé les traits mentaux et ceux du tempérament pour découvrir l'étendue du choix d'époux comme déterminé par les coefficients de corrélation, et les résultats comparés à ceux pour un groupe de contrôle formé par un système de choix au hasard. Dans les traits physiques, on a analysé la relation entre les époux et les pères de leurs femmes et celle entre les épouses et les mères de leurs maris pour tester certaines théories de Freud.

On a trouvé que dans tels traits physiques que l'âge, la hauteur, le poids, la couleur des cheveux et des yeux, et dans la relation des beaux-parents dans ces traits, et dans tels traits mentaux que les réactions d'association, la capacité de raisonner dans le calcul, la portée générale de ses connaissances, et dans les opinions sur les sujets de l'intérêt quotidien, les coefficients de corrélation par leur grandeur et leur caractère positif indiquent une tendance à l'homogamie. L'absence de corrélations positives ou négatives dans le test de vocabulaire et dans les deux tests d'émotionnalité employés indiquent un choix d'époux au hasard à ces égards. Dans l'éducation, le métier, et la préférence politique une tendance à l'homogamie est encore indiquée.

Sauf à ces trois derniers égards et dans le vocabulaire, les coefficients de corrélation pour le groupe de contrôle de choix au hasard sont constamment insignifiants. Ceux obtenus dans cette étude s'accordent dans leurs tendances avec les recherches antérieures dans ce domaine.

SCHILLER

EINE QUANTITATIVE ANALYSE DER HEIRATSWAHL IN EINER KLEINEN GRUPPE

(Referat)

Von einer Gruppe bestehend aus 46 verheirateten Paaren, die in Bezug auf Alter, Erziehung, Religion, Beruf, und sozial-ökonomischen Stand homogen waren, wurden Daten gesammelt über gewisse physische, geistige, und Temperaments-Eigenschaften. Man analysierte dann diese Befunde um den Umfang der Selektivität bei der Paarung (extent of selective mating) mittels Korrelationskoeffizienten zu erforschen, und verglich dann die Befunde mit denen, die an einer Kontrollgruppe erzielt worden waren, die mittels eines Systems der zufälligen Paarung gebildet worden war. In Bezug auf die physischen Eigenschaften analysierte man u.a. auch die Beziehung der Gatten zu den Vätern der Gattinen und die Beziehung der Gattinen zu den Müttern der Gatten, um gewisse Theorien von Freud zu prüfen.

Bei physischen Eigenschaften wie z.B. Alter, Grösse, Gewicht, Haarfarbe und Augenfarbe, bei der Beziehung zu den Schwiegereltern in Bezug auf diese Eigenschaften, bei geistigen Eigenschaften wie z.B. Assoziationsreaktionen, Fähigkeit zu mathematischem Denken, und allgemeine Weite der Kenntnisse, und bei Meinungen über Gegenstände laufender Interesse, erwiesen die Korrelationskoeffizienten durch ihre Grösse und ihre positive Richtung eine Neigung zur Homogamie. Der Mangel sowohl an positiver wie an negativer Korrelation bei der Wortschatzprüfung (vocabulary test) und bei den zwei Prüfungen der Affektivität (emotionality) die gebraucht wurden weist auf Zufälligkeit der Paarung in diesen Bezügen hin. Bei Erziehung, Beruf, und politischer Bevorzugung wird wieder eine Neigung in die Richtung der Homogamie angedeutet.

Mit Ausnahme dieser drei letzten Beziehungen und des Wortschatzes erwiesen sich die Korrelationskoeffizienten bei der "zufälligen" Kontrollgruppe als übereinstimmend unbedeutend. Die in dieser Untersuchung erhaltenen Befunden stimmen in ihrer allgemeinen Richtung mit den Befunden aus früheren Untersuchungen überein.

SCHILLER

THE RELATION BETWEEN LIBERAL AND CONSERVATIVE ATTITUDES IN COLLEGE STUDENTS, AND OTHER FACTORS*

From the Division of Educational Reference of Purdue University

A. J. HARRIS, H. H. REMMERS, AND C. E. ELLISON

In 1922 Harper (3) gave a liberalism-conservatism questionnaire to 2900 educators, from all parts of the country. Their education varied from not more than four years above the eighth grade to doctorate work. From a study of internal inconsistencies, tendency to agree with a statement as made, etc., he came to the following conclusions: "Other things being equal, the greater the conservatism the less stable, the less independent, the less consistent, the less scientific in attitude is the thinking on the problems involved in the test. Conversely, it has appeared that the greater the non-conservatism, the greater the resources adapted to reflective reaction to the more primary and essential social conceptions and issues. It would seem that conservatism is characterized not by cautious and alert inquiry but by dangerous and satisfied blindness."

In addition to these general considerations, he found:

1. A slight difference, not statistically reliable, between the sexes, the men being more liberal;
2. No significant difference between Catholics and Protestants, or between the Protestant denominations;
3. No significant difference between the two major political parties;
4. A slight tendency towards higher liberalism with increasing age, for educators with college degrees;
5. An r of $.521 \pm .032$ between liberalism and amount of education;
6. A decided rise in liberalism in those taking certain thought-provoking courses in social and educational problems.

Symonds (4) tested students from the eighth grade in Honolulu public schools through the University of Hawaii, with his "Social

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, July 2, 1931.

Attitudes Questionnaire." He found practically no relation between amount of education and liberalism, an r of .28 between an information test and liberalism, and an r of .28 between intelligence and liberalism. He also found that men were slightly more liberal than women. In commenting on these results, he said: "What the main factors are leading to the formation of attitudes we do not know; we surmise, however, that more important than the intellectual factors are the emotional."

Allport (1) gave a questionnaire to 375 elementary psychology students at Dartmouth, one month before the presidential election of 1928. In the questionnaire were items to reveal the student's radicalism or conservatism, his knowledge of the issues of the campaign, prejudices, choice of candidate, religious affiliation, father's political preference, etc. In comparing the 10 per cent most radical with the 10 per cent most conservative, he obtained the following results:

1. Radicals showed much less prejudice than conservatives.
2. Radicals had no more correct information, but much less misinformation than the conservatives.
3. Radicals had higher grades than the conservatives.
4. Radicals disagreed with their fathers' political opinions far more frequently.
5. Radicals tended to be less religious.
6. Radicals felt less strongly about politics.

The studies of Harper, Symonds, and Allport found interesting relationships between liberalism-conservatism and several other factors. It was our purpose, in the present study, to check these relationships and to investigate the presence of other relationships.

PLAN OF THE PRESENT STUDY

Harper's "Social Study" and a questionnaire were given to the members of the elementary psychology course at Purdue University, chiefly sophomores. The Harper test consists of 71 statements, to be marked in agreement or disagreement. In this study no further criterion of liberalism than this was set up. A high score on it shows liberalism, a low score, conservatism. The questionnaire follows:

Do not put your name on this sheet

1. Place a circle around the figure that represents where you belong, with respect to your views on political, social, economic, and religious issues:
- | | | | | | | |
|--------------|---|---|---------|---|---|---------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| extreme | | | average | | | extreme |
| conservative | | | liberal | | | liberal |
2. Age to nearest birthday 3. Sex
4. To what church do you belong?
5. If not a member of any church, what church do you prefer?
6. With what church is your father affiliated?
7. With what church is your mother affiliated?
8. How often do you attend church (underline which):
(a) regularly; (b) fairly regularly; (c) seldom; (d) never
9. Has there been any marked change in your attitude toward religion since you entered high school? Yes No (underline which)
10. Was this change in the nature of: (underline which)
(a) greater faith
(b) less faith
(c) change in affiliation
11. Do you believe that the story of the creation of the world given in the Bible is literally true? (underline which) Yes No
12. Do you believe that man has evolved from a lower type of animal? (underline which) Yes No
13. What is your father's occupation? (If he is dead, state what his occupation was):
14. What is the population of the community in which you have spent most of your life? (underline which):
(a) rural; (b) town of less than 20,000; (c) 20,000 to 50,000;
(d) 50,000 to 100,000; (e) above 100,000.
15. To what political party or parties do your parents belong? Indicate by placing an *F* for 'father' and an *M* for 'mother' before the name of the proper parties.
..... Independent Socialist Farmer-Labor
..... Democratic Republican
16. To what political party do you belong?
..... Independent Socialist Farmer-Labor
..... Democratic Republican
17. What percentage of your college expenses are you earning?

The students' intelligence rating was obtained from the American Council on Education Psychological Examination given to all freshmen during orientation week.

TABLE 1
LIBERALISM OF MEN AND WOMEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{Diff.}{\sigma\ diff.}$
Men	160	40.03	8.01	1.83	2.08
Women	147	38.20	7.44		

The students were instructed not to sign their names either to the questionnaire or the test. When these forms had been completed, however, the students were asked to sign them for purposes of investigation. They were assured that their markings would be treated as strictly confidential. Relatively, only a few failed to sign the papers; a total of 307 out of 342 being signed. The two sexes were approximately equally represented.

RELIABILITY

Harper found an r of $.782 \pm .017$ between alternate halves of the test taken by a random selection of 170 graduate teachers. In the present study by the same method we found for a random sampling of 100 cases a reliability of .774, which closely corresponds to Harper's results. No reliability measures were obtained for the questionnaire data.

RESULTS

The mean for the entire group of 307 students was 39.16 with a *S.D.* of 7.74. These scores ranged from 16 to 65. This mean corresponds substantially to Harper's median scaled score of 47 for educators holding a bachelor's degree only in the middle and far West. There were 98 chances in 100¹ of a true difference between men and women.

This difference between the sexes is about the same as that found by Symonds and that found by Harper between the average scores of female graduate educators and male graduate educators.

Intelligence and Liberalism. The r between liberalism scores and percentile scores on the American Council on Education Psychological Examination for 149 men was found to be $.29 \pm .05$, and for 139 women, $.09 \pm .05$. This corresponds very closely to the correlation of .28 found by Symonds between intelligence and liberalism and to the correlation of .21 between radicalism and college grades reported by Allport. There is evidently some relationship here, although the slight degree of correlation shows that factors other than intelligence are of most importance in determining a person's liberalism or conservatism.

¹The number of chances in 100 of a true difference are taken from Table XIV in "Statistics in Psychology and Education" by H. E. Garrett (2), which is based on the theory that a difference which is three times its *S.D.* indicates that there is a true difference greater than zero.

Self-Estimates of Liberalism. In some of the psychology classes the students were asked to estimate their own liberalism soon after being told their scores on the Harper test. The rest of the students were asked to make their estimate before being told their scores on the Harper test. The correlations for the groups are given below. Group 1 is that group of students who estimated their liberalism before being told their scores on the Harper test.

These results tend to show that the students' estimates of their own liberalism are not very accurate. There is a suggestion that knowledge of their liberalism scores had a greater effect upon the estimates of the men than upon those of the women. The differences are not in any case reliable.

Church Attendance and Liberalism. An r of $-.275 \pm .051$ was obtained between regularity of church attendance and Harper scores for 146 men. For 130 women the r was $-.125 \pm .059$. An r (bi-serial) between liberalism and church attendance was also obtained using only the extreme groups: that is, those who attend church regularly and those who never attend.

TABLE 2
LIBERALISM SCORES AND SELF-ESTIMATES OF LIBERALISM

	r	P.E.	N
Group I men	.317	.070	80
Group I women	.342	.094	36
Group II men	.491	.073	56
Group II women	.398	.057	95

TABLE 3
CHURCH ATTENDANCE AND LIBERALISM, EXTREME GROUPS

	Regular	Never	r (bis)	P.E.
Men	24	11	-.646	.101
Men and women	57	13	-.639	.084

No r was obtained for the women alone since only two of them never attended church. From the above results we may say that those who never attend church tend to be much more liberal than regular church-goers.

Comparison of Religious Denominations. Table 4 gives the means for the various religious denominations, for those who have no

church preference, and for the combined Protestant denominations. Samplings of other denominations were too small to allow comparison. Our findings correspond to Harper's in that we found no significant differences between the denominations. Those having no church preference, however, have a mean which is 6.47 above that of the combined Protestants. This difference is completely reliable, being 4.47 times its *S.D.*, indicating that the average student who has no church preference is decidedly more liberal than those who have some religious affiliation.

Belief in Evolution and Liberalism. Reliable differences were found between those who believed in evolution but not in the story

TABLE 4
MEANS OF RELIGIOUS DENOMINATIONS, COMBINED PROTESTANTS,
AND NO PREFERENCE

	<i>N</i>	<i>M</i>	σ
No preference	36	45.32	7.64
Combined Protestants	220	38.96	7.67
Methodist	80	39.69	8.85
Baptist	13	38.85	7.58
Episcopalian	10	38.80	6.01
Christian Science	14	38.64	4.93
Lutheran	11	38.09	7.36
Presbyterian	33	38.00	7.29
Christian	16	36.44	6.40
Catholic	18	36.67	6.07

TABLE 5
BELIEF IN EVOLUTION AND LIBERALISM: MEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
Belief in evolution but not in biblical creation story	72	44.33	8.01	9.54	7.7
Belief in biblical creation story, but not in evolution	41	34.79	5.10		

of creation literally as given in the Bible, and those who believed the biblical creation story but did not believe in evolution.

This result seems to show that a belief in evolution is a fairly good index of liberalism for men.

TABLE 6
BELIEF IN EVOLUTION AND LIBERALISM: WOMEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{diff.}}$
Belief in evolution but not in biblical creation story	47	39.70	9.17	3.45	2.09
Belief in biblical creation story, but not in evolution	36	36.25	5.80		

There are 98 chances in 100 of a true difference here.

For women the difference is appreciably smaller than for men. It is noteworthy that 65 per cent of the men believe in evolution, as against 57 per cent of the women. This difference is statistically significant.

It was found that those men who believed neither in evolution nor in the literal story of the creation of the world as given in the Bible had an average score below that of those who believed in evolution and above those who believed in the Bible story, which probably indicates that they are in a transition stage in forming their opinions. This group of 19 men had a mean of 39.95 with a *S.D.* of 6.00.

TABLE 7
DIFFERENCES: MEN

	Belief in evolution, disbelief in Bible story	Belief in neither
Belief in neither	↑ 4.38*	
Belief in Bible Story, disbelief in evolution	↑ 9.54*	↑ 5.16*

*Indicates 98 chances in 100 or better, of a true difference. The arrows indicate the category having the advantage of the difference.

Eleven men marked both items "yes." This inconsistency was probably due to a misunderstanding of the questions.

Eight women marked both items "no." These women had a mean of 43.50. It is unfortunate that there were no more cases to allow comparisons to be made. Only two women marked both items "yes."

Change in Belief in Religion. The questionnaire obtained information as to whether the students had had any change in belief in religion since they entered high school. There were 91 chances in 100 of a true difference between those men who changed in the direction of less faith and those who gained faith.

TABLE 8
CHANGE IN FAITH AND LIBERALISM: MEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
More faith	20	39.75	6.72	2.56	1.34
Less faith	49	42.31	8.37		

This seems to suggest that there is a tendency toward an inverse relationship between religious faith and liberalism.

For women there are 79 chances in 100 of a true difference.

TABLE 9
CHANGE IN FAITH AND LIBERALISM: WOMEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
More faith	29	39.45	5.91	1.50	.794
Less faith	30	40.95	8.46		

There were 99 chances in 100 of a true difference between a select group of men who had had no change in belief since entering high school and no religious affiliations, and those who had had no change in belief, but who had a church preference.

TABLE 10
HAVING CHURCH PREFERENCE AND LIBERALISM: MEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
No preference—no change	11	43.55	5.52	4.73	2.42
Church preference—no change	51	38.82	7.34		

There were only three women who had had no church preference since entering high school. These three had scores of 56, 45, and 42. It seems reasonable to assume that a large enough group of women in this classification would have been significantly high in liberalism. No significant differences were found between those who changed from one faith to another and those retaining the same faith.

Political Affiliations and Liberalism. Our results show no systematic difference between the two major parties. For the men the Democrats were slightly more liberal, but for the women the Republicans had a slightly higher score. For both sexes the Independents were more liberal than the members of either party.

TABLE 11
POLITICAL AFFILIATION AND LIBERALISM: MEN

	<i>N</i>	<i>M</i>	σ
Independents	70	42.64	9.06
Republicans	41	38.12	6.63
Democrats	29	40.45	6.48
<i>Differences</i>			
	Independents	Republicans	
Republicans	↑ 4.52*		
Democrats	↑ 2.19	←1.59	

*To be read same as in Table 7.

TABLE 12
POLITICAL AFFILIATION AND LIBERALISM: WOMEN

	<i>N</i>	<i>M</i>	σ
Independents	63	40.33	7.56
Republicans	48	37.25	7.17
Democrats	19	34.74	5.37
<i>Differences</i>			
	Independents	Republicans	
Republicans	↑ 3.08*		
Democrats	↑ 5.60*	↑ 2.51	

*To be read same as in Table 7.

It is interesting to note that 50 per cent of the men and about 48 per cent of the women indicated that they were independent politically. This high proportion is, of course, in part due to the fact that many have as yet found no interest in politics either because they are not old enough to vote or because they are lacking in civic interest. It seems logical to assume that these students materially reduced the mean for the Independent group.

There is no significant difference between those men whose parents agree on politics and those whose parents disagree.

TABLE 13
POLITICAL AGREEMENT OF PARENTS AND LIBERALISM: MEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
Parents disagree	25	40.32	6.29		
Parents agree	95	40.17	7.98	.15	.10

For women there are 99 chances in 100 of a true difference.

TABLE 14
POLITICAL AGREEMENT OF PARENTS AND LIBERALISM: WOMEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
Parents disagree	20	41.85	6.43		
Parents agree	94	37.66	7.72	4.19	2.55

It is of interest to note the large number of parents who agree on politics; 79 per cent of the parents of the men and 82 per cent of those of the women. It is evident that people tend to choose mates who agree with them on politics, or that either the husband or wife, in most marriages, dominates over any notions about politics that the other may have had before marriage.

There are 96 chances in 100 of a true difference between those men who disagree with their fathers' politics and those who agree.

TABLE 15
AGREEMENT WITH FATHER'S POLITICS AND LIBERALISM: MEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
Disagree	70	41.19	7.74		
Agree	59	38.95	6.63	2.24	1.82

In the case of the women there is an almost completely reliable difference.

TABLE 16
AGREEMENT WITH FATHER'S POLITICS AND LIBERALISM: WOMEN

	<i>N</i>	<i>M</i>	σ	<i>Diff.</i>	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
Disagree	57	40.42	7.89	3.92	2.93
Agree	66	36.50	6.81		

The greater difference for the women than for the men suggests that women are more liberal when they disagree with their fathers' politics than are men.

Besides the preceding categories there was a select group of 12 men who were independent politically, and whose parents were also independent. There were 96 chances in 100 of a true difference between these men and all the men who were independent, 98.6 chances in 100 of a true difference between these men and the Democrats, and a completely reliable difference between these men and the Republicans. The small number here, of course, makes for a very large sampling error.

TABLE 17
DIFFERENCES: MEN

	Student and parents Independents	Student Independents	Democrats
Students	↑		
Independents	5.37		
	↑		
Democrats	6.81*	2.19	
	↑	↑	
Republicans	9.88*	4.52*	↑ 1.59

*To be read as in Table 7.

These findings indicate a tendency for those who are independent politically, and especially those whose parents are also independent, to be liberal. It seems that a family where neither parent has any political affiliation tends to be liberal, and, therefore, a student coming from such a family tends not to have a political preference and to be liberal.

Size of Home Community and Liberalism. An r was computed between population of the students' home towns and their liberalism scores. For 144 men this r was $+.029$, and for 131 women it was $-.009$; both substantially zero.

Self-Support and Liberalism. An r was also obtained between the percentage of college expenses earned by the students and their liberalism scores. These were insignificant for both men and women: for men, $-.117 \pm .054$, and for women, $-.0008$.

Father's Occupation and Liberalism. The occupations of fathers as reported by the students were classified in six groups, professional, business and technical, skilled labor, semi-skilled labor, unskilled labor, and agriculture. There were no representatives of the unskilled group, and only five in all representing the semi-skilled group. The relationships between liberalism and fathers' occupations are shown in Table 18 for men, and in Table 19 for women. The semi-skilled group, while very scantily represented, gives indications that with more representatives it would be the least liberal. None of the differences between the four groups that are fairly well represented are statistically reliable, but they suggest interesting relationships. The men whose fathers are in business and technical occupations are the least liberal; presumably this should be the class most satisfied with things as they are. For the women, fathers' occupations in descending order of liberalism are: Professional,

TABLE 18
FATHER'S OCCUPATION AND LIBERALISM: MEN

Father's occupation	N	M	σ
Professional	17	41.64	7.32
Business and technical	57	39.42	7.56
Skilled labor	38	42.09	8.01
Semi-skilled labor	2	36.00	
Agriculture	22	40.50	7.72

<i>Differences</i>			
	Professional	Business and technical	Skilled labor
Business and technical	↑ 2.22		
Skilled labor	← .45	← 2.67	
Agriculture	↑ 1.14	← 1.08	↑ 1.59

TABLE 19
FATHER'S OCCUPATION AND LIBERALISM: WOMEN

Father's occupation	N	M	σ
Professional	18	40.50	8.01
Business and technical	52	39.10	8.64
Skilled labor	19	38.70	7.65
Semi-skilled labor	3	29.00	
Agriculture	36	37.08	5.01

Differences

	Professional	Business and technical	Skilled labor
Business and technical	↑ 1.38		
	↑	↑	
Skilled labor	1.81	.43	
	↑	↑	↑
Agriculture	3.42	2.04	1.61

business and technical, skilled labor, agriculture. It may be that here the difference is one of cultural opportunity.

Effect of a Certain Type of Education on Liberalism. The Harper test was given to a class of 23 juniors and seniors nearing the end of a course in sociology. A very definite difference was found between these students and those from the elementary psychology course.

TABLE 20

	N	M	σ	Diff.	$\frac{\text{Diff.}}{\sigma \text{ diff.}}$
Sociology class	23	50.61	9.30		
				11.45	5.58
Elementary psychology classes	307	39.16	7.74		

The range of scores in the elementary psychology classes was 16-65; in the sociology class it was 37-68. The higher scores of the sociology students may have been the result of training in the course or of selection, or of both. Harper's results showed a very marked increase of liberalism in a group of graduate educators as a result of taking a somewhat similar course. The problem is worthy of further study.

SUMMARY AND GENERAL CONCLUSIONS

Suggestive relationships and lack of relationships were found between factors and liberalism.

1. *Liberalism of the Group as a Whole.* The mean score for 307 students was 39.16, which corresponds to Harper's scaled median score of 47 for educators holding a bachelor's degree only, in the middle and far West. This seems to indicate that as a group Purdue sophomores are, relatively, fairly liberal. Their general viewpoint is anti-socialistic.

2. *Relation to Intelligence.* The r found between intelligence and liberalism for men was $.29 \pm .051$. For women it was only $.09 \pm .055$. Allport found an r of .21 between radicalism and college grades. Symonds found a correlation of .28 between intelligence and liberalism. These results all seem to show a slight relationship between intelligence and liberalism and suggest that other factors than intelligence are of major importance in creating a liberal attitude.

3. *Self-Estimates of Liberalism.* There seems to be a slight relation between self-estimates of liberalism and liberalism scores, but a student's estimate of his own liberalism would be unreliable as an index of liberalism. There is no significant difference in the ability of the two sexes in this respect.

4. *Religion and Liberalism*

1. Liberals tend to be less religious, this tendency being stronger in men than in women.

2. Those who never attend church are more liberal than those who attend at all, and much more liberal than those who attend regularly.

3. There are no significant differences between the various religious denominations which were represented in the study.

4. Those having no church preference are decidedly more liberal than those having a preference for some church.

5. Men who believe in evolution are decidedly more liberal than those who do not. Women exhibit the same tendency, but not so strongly.

6. Those men who believe neither in evolution nor in literal interpretation of the Bible story of creation have a mean score between those who believe in evolution and those who do not, probably indicating a transition stage in the formation of their opinions.

7. Those men who have lost some degree of faith in religion since entering high school are more liberal than those who have gained faith. Women show this tendency even more strongly.

8. Men who have had no religious affiliation since entering high school are more liberal than those who have had the same religious affiliation and no change in faith since entering high school.

9. Change from one denomination to another seems to have practically no relation to liberalism.

5. *Political Affiliations and Liberalism*

1. There are no systematic differences between the two major parties in regard to liberalism, the Democrats being slightly higher among the men, and the Republicans among the women.

2. Those who have no political preference are more liberal than those who have. The difference is larger for the women than for the men.

3. Between those whose parents disagree and those who agree on politics there is no difference for the men, but a significant difference for the women.

4. Those who disagree with their fathers' politics tend to be more liberal. The women show a greater difference in this respect than do the men. Allport found this same tendency in his study.

5. Those who are independent politically, and whose parents are also independent, show a tendency to be highly liberal.

6. *Size of Home Community.* There is practically no correlation between the size of home town and liberalism.

7. *Percentage of College Expenses Earned.* There is practically no correlation between percentage of expenses earned and liberalism.

8. *Father's Occupation and Liberalism.* Suggestive but unreliable differences in liberalism were found when the students were grouped according to the occupation of their fathers.

9. *Education and Liberalism.* A class in sociology was found to be much more liberal than the elementary psychology class. This suggests once again the profound importance of school training in moulding public opinion, and confirms Harper's findings on the importance of certain kinds of education.

10. *Sex Differences.* The following differences were found:

a. Men are slightly more liberal.

b. Only two women out of 130 never attend church.

c. There is a lower r between belief in evolution and liberalism for women than for men.

d. There is a lower r between intelligence and liberalism for women than for men.

e. There is a significant difference between women whose parents disagree on politics and those whose parents agree, and none for the men.

f. There is a greater difference between those women who disagree with their fathers' politics and those who do not than there is for the same categories among men.

All this points to a tendency for women to be governed by parental and public opinion to a much greater extent than are men. This tendency is not manifested much in their ideas, since they are only slightly less liberal than men; but it is shown in their lack of outward manifestation of liberalism, in the ways itemized above.

Purdue liberals as a group tend to be more intelligent, to be less religious, to believe more in evolution, to be more independent politically than Purdue students as a whole, who, in turn, exhibit these tendencies more than the conservative students.

REFERENCES

1. ALLPORT, G. W. The composition of political attitudes. *Amer. J. Sociol.*, 1929, **35**, 220-238.
2. GARRETT, H. E. Statistics in psychology and education. New York: Longmans, Green, 1927. Pp. 317.
3. HARPER, M. H. Social beliefs and attitudes of American educators. *Teach. Coll. Contrib. Educ.*, 1927, No. 294. Pp. 91.
4. SYMONDS, P. M. A social attitude questionnaire. *J. Educ. Psychol.*, 1925, **16**, 316-322.

Purdue University
Lafayette, Indiana

LA RELATION ENTRE LES ATTITUDES LIBÉRALE ET CONSERVATRICE PARMI LES ÉTUDIANTS UNIVERSITAIRES ET D'AUTRES FACTEURS

(Résumé)

On a fait subir "l'Etude Sociale" de Harper et un questionnaire supplémentaire de renseignements personnels à 342 étudiants de la psychologie élémentaire à l'Université de Purdue. En employant les résultats du Test Harper comme mesure de libéralisme-conservatisme, on a étudié les relations possible entre le libéralisme-conservatisme et les facteurs suivants: le sexe, l'estimation par soi-même de son libéralisme, la régularité de l'assistance à l'église, la préférence de l'église, la croyance ou le manque de croyance à l'évolution, l'affiliation politique, l'affiliation politique des par-

ents, la grandeur de la ville natale, le montant de travail nécessaire pour gagner sa propre vie à l'université, et le métier du père. Les estimations de leur libéralisme par les étudiants n'ont pas été très précises. Les hommes ont été un peu plus libéraux que les femmes. Il y a eu des indications que les femmes ont été influencées plus que les hommes par l'opinion des parents et l'opinion publique. Les libéraux comme groupe tendaient à être plus intelligents, à être moins religieux, à croire plus à l'évolution, et à être plus indépendants dans la politique que l'ensemble des étudiants, qui ont montré à leur tour ces tendances plus que les étudiants conservateurs. Une classe étudiant la sociologie s'est montrée beaucoup plus libérale que le groupe des étudiants de la psychologie élémentaire.

HARRIS, REMMERS, ET ELLISON

FREISINNIGE UND KONSERVATIVE EINSTELLUNGEN VON STUDENTEN IM ZUSAMMENHANG MIT ANDEREN EINWIRKUNGEN

(Referat)

Es wurden 342 Studenten aus einem Anfangsunterricht in der Psychologie an der Purdue University der sozialen Untersuchung (Social Study) von Harper und einer ergänzenden Fragenliste über persönliche Gelegenheiten (personal information questionnaire) unterworfen. Man gebrauchte die an der Harper Social Study erzielte Zahl (score) als Massstab des Liberalismus-Konservatismus und forschte dann nach möglichen Beziehungen zwischen Liberalismus-Konservatismus und folgenden Einwirkungen: Geschlecht, Selbst-Abschätzung des eigenen Liberalismus, Regelmässigkeit des Kirchengesuchs, bevorzugte Kirche, Glaube oder Unglaube der Entwicklungstheorie (evolution) gegenüber, politische Verbindung, Grösse der Heimatgemeinschaft, Grad der Selbsterhaltung, und Beruf des Vaters. Die Abschätzungen des eignen Liberalismus waren nicht sehr genau. Die Männer zeigten sich als in ganz geringem Masse freisinniger als die Frauen. Man fand Andeutungen, dass die Frauen durch die Einwirkung elterlicher und öffentlicher Meinungen stärker beeinflusst wurden als die Männer. Die Freisinnigen neigten im grossen Ganzen dazu, intelligenter und weniger religiös zu sein, stärker an die Evolution zu glauben, und politisch selbständiger zu sein als die Studenten im Allgemeinen. Letzterwiesen ihrerseits diese Eigenschaften in stärkerem Grade als die konservativen Studenten. Eine Klasse die Soziologie studierte erwies sich als viel freisinniger als die Anfangsklasse in der Psychologie.

HARRIS, REMMERS, UND ELLISON

SOME ASPECTS OF TABOO AND TOTEMISM*

HOWARD BECKER AND DAVID K. BRUNER

The tentative hypothesis of the origins and development of taboo and totemism outlined in the present paper¹ employs the principles of ambivalence, projection, and displacement—principles which social psychology owes to Freud. It diverges sharply, however, from the theories of Freud as set forth in *Totem and Taboo*, for the latter have been rendered untenable by inconsistencies brought to view by recent psychiatric discoveries and ethnological research. After noting one or two fundamental postulates, and defining the terms to be used, we shall summarize Freud's theories and the criticisms offered by Kroeber, Malinowski, and Mead, and shall then propose an alternative hypothesis which, although avowedly conjectural, seems somewhat more tenable.

I

Our basic postulates, chosen in the light of present anthropological knowledge, are that (1) there is no essential difference between preliterate thinking and that of the "modern" mind, and (2) that the preliterate mind functions in essentially the same fashion wherever found, i.e., with fundamentally identical psychological processes underlying a wide diversity of culture patterns. The alleged "pre-logical" character of preliterate thinking is the result of animatistic, animistic, and other explanatory categories developed by the preliterate in the absence of the contradicting facts which have accrued to the modern group during centuries of "superior" culture. Recent observers state that where cause and effect are apparent, as in the case of house construction or other material-using activities of everyday life, the preliterate acts in the same way and with as much insight as does the "modern" man in similar situations. Furthermore, the "primitive-civilized" contrast set up by Lévy-Bruhl and his school holds only for preliterate man, on the one

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, May 29, 1931.

¹The essentials of the hypothesis were set forth in 1926 in an unpublished M.A. thesis, "A Social-Psychological Study of Bereavement" (Northwestern University), by the first-mentioned of the writers of the present article.

hand, and the scientist at work in his own field, on the other: "civilized" men, and even "Ph.D.'s at play" present plentiful examples of "prelogical" thinking.

II

Definition of the term "taboo" is next in order. "Our own word, 'taboo,' is derived from the Polynesian 'tapu' . . . but [the Anglicized word] is far from designating all the ramifications and shades of meaning that clustered about its prototype," says Lowie (5, p. 78). "The aboriginal term was, like *mana*, a highly abstract concept, but while *mana* invariably denoted a startling deviation from the norm, *tapu* was distinctly ambivalent: it represented the holiness of the divine, setting it off from the profaneness of *noa* (or common 'things') . . . and persons; but that very attitude which lent mystic power to the dieties and their earthly vice-regents invested them with a 'mysterious perilousness and unapproachableness' for the uninitiated, or indeed, for any one of lesser quality."

Equivalents of the word "taboo" may be found in the languages of other races, and the significant point about such words is that they have the peculiar ambivalence noted in the definition quoted. *Kodesh*, the Hebrew equivalent of our English word "holy," had the double meaning of "holy" and "unclean"; the Latin *sacer* from which our "sacred" is derived could mean either "blessed" or "cursed." It was supposed by earlier anthropologists that originally the word "taboo" did not mean both holy and unclean but signified something weird and demonic, something which might not be touched, thus emphasizing a characteristic common to both extremes of the later conception. This persistent common trait seems to show, however, that an original correspondence existed between what was holy and what was unclean, which only later became differentiated. The double meaning in question may thus have belonged to the word "taboo" from the very beginning; it now serves to designate a definite ambivalence as well as everything which has come into existence on the basis of this ambivalence. "Taboo" is itself an ambivalent word, and even the established meaning might of itself have allowed us to guess what later schools of ethnology have, as the result of extensive investigation, come to regard as probable—namely, that the taboo prohibition may sometimes be explained as the result of an emotional ambivalence.

Now *abnormal* emotional ambivalence need be present in only

one or two individuals of a preliterate group in order to affect its customs and ceremonials (as will be indicated later), but with this reservation the words of Freud have unusual significance at this point:

"Taboo is a very primitive prohibition imposed from without (by an authority) and directed against the strongest desires of man. The desire to violate it continues in the unconscious; persons who obey the taboo have an ambivalent feeling toward what is affected by the taboo. The magic power attributed to the taboo goes back to its ability to lead man into temptation; it behaves like a contagion, because the example is contagious, and because the prohibited desire becomes displacing in the unconscious upon something else. The expiation for the violation of a taboo through a renunciation proves that a renunciation is at the basis of the taboo" (2, p. 59).

III

"Totemism" has been variously defined, but in general it may be said that it is a system of sib organization which involves the worship or reverencing of some organism which is considered the ancestor of the sib. In addition, members of a totem are forbidden to kill the totem or to eat it, or if they do, they must placate the totem in some way. Further, members of a totem-group quite frequently are forbidden to have sexual intercourse with each other. The totem may be a plant rather than an animal, and there are almost endless variations in this and other details (e.g., totemic groups are sometimes endogamous); but the usual features are as stated. There is much difference of opinion as to the interdependence of the several points involved in the above definition: whether, for instance, the religious and social-organizational aspects of totemism have always coexisted or have had independent origins. Leaving this, as well as the problem of the origin of the complicated system of marriage classes found in Australian tribes and elsewhere, aside for the present, it is sufficient here to note that great care is usually taken to prevent incest. Not only is the totemic system often designed, to all appearances, with this purpose chiefly in view, but there is in addition a series of customs guarding the individual's behavior toward near relatives (biological and/or classificatory), which may be called avoidances or taboos. These are maintained with almost religious severity, and there can hardly be any doubt that their object is to prevent incest. They are found in various parts of

the world, and seem to call for a psychological rather than a cultural explanation.

IV

Ambivalence means that hostile and affectionate feelings are directed toward one and the same object of interest. It is conspicuous only in definitely neuropathic or psychopathic personalities, although some psychiatrists say that there is a streak of ambivalence in all our personal relations, whether consciously dominated by tender feeling or by the opposite feeling of hatred. Since one or two neuropaths or psychopaths in a community are enough, if in a position of influence, to set up a culture-pattern for that community, marked ambivalence in "normal" persons need not be assumed. It is well established, moreover, that among preliterate peoples neuropathic and psychopathic types often find socially approved outlets for their behavior in shamanism, etc.

As a result of ambivalence, either dominant hatred or tender feeling may lead to identical outward behavior. Thus a man with an Oedipus complex may compensate for his repressed mother-love by constant quarreling and other hate-behavior during his mother's life; but at her death convention sanctions expression of the repressed love, which may then be manifested in exaggerated mourning. The repressed death-wish of a wife, on the other hand, consciously compensated for by excessive attentions while her husband lives, is even more sternly repressed by the demands of convention at his death and, as in the first case, may result in mourning behavior that passes all "normal" bounds.

Projection offers an outlet for the socially unacceptable hatred aspect of ambivalent emotions by attributing the felt hostility to the object. Offering projection as an explanation of fears of the dead, Freud thus describes the process: "This unknown hostility, of which we are ignorant and of which we do not wish to know, is projected from our inner perception into the outer world and is thereby detached from our own person and attributed to the other. Not we, the survivors, rejoice because we are rid of the deceased, on the contrary, we mourn for him; but now, curiously enough, he . . . would rejoice in our misfortune and . . . seeks our death. The survivors must now defend themselves against this; . . . they are freed from inner oppression, but they have only succeeded in exchanging it for an affliction from without" (2, p. 106).

Displacement or substitution is a third process; its manifestations have been abundantly verified by psychiatric research. The ambivalent or other feeling toward an object may be transferred to another object which serves as a symbol of the first; in Freud's theory of totemism and taboo the transfer is from the father to the totem animal.

Having thus briefly defined the terms used, we turn to Freud's theory as stated in *Totem and Taboo*.

He commences with the inference of Darwin, developed further by Atkinson, that at a very early period man lived in small communities consisting of an adult male and a number of females and immature individuals, the males among the latter being driven off by the head of the group as they became old enough to evoke his jealousy. To this Freud adds the Robertson Smith theory that altar sacrifice is the essential element in every ancient cult and that such sacrifice goes back to a killing and eating by the sib of its totem animal, which was regarded as of kin with the sib and as its god, and whose killing at ordinary times was strictly forbidden. The Oedipus complex² directed upon these two hypotheses welds them into an instrument with which it is possible for the Freudian to dissect and "explain" most of the essentials of human civilization, as follows:

The expelled sons of the primal horde finally banded together, slew their father, ate him, and appropriated the females. In this they satisfied the same hate impulse that is a normal infantile trait and the basis of most neuroses, but which often leads to "unconscious displacement" of feelings, especially upon animals. At this point, however, the ambivalence of emotions proved decisive. The tender feeling which had always persisted by the side of the brothers' hate for their father gained the upper hand as soon as this hate was satisfied, and took the form of remorse and sense of guilt. "What the father's presence had formerly prevented they themselves now prohibited in the psychic situation of 'subsequent obedience' which we know so well from psychoanalysis. They undid their deed by declaring the killing of the father substitute, the totem, was not

²"If the totem animal is the father, then the two main commandments of totemism, the two taboo rules which constitute its nucleus,—not to kill the totem animal and not to use a woman belonging to the same totem for sexual purposes,—agree in content with the two crimes of Oedipus, who slew his father and took his mother to wife" Freud *op. cit.*, p. 219.

allowed, and renounced the fruits of their deeds by denying themselves the liberated women. Thus they created the two fundamental taboos of totemism." These are "the oldest and most important taboos" of mankind; namely, not to kill the totem animal and to avoid sexual intercourse with totem companions of the opposite sex, alongside of which many if not all other taboos are "secondary, displaced and distorted." The renunciation of the women or incest prohibition had also this practical foundation: that any attempt to divide the spoils, when each member of the band really wished to emulate the father and possess all the women, would have disrupted the organization which had made the brothers strong (2, p. 236). The totem sacrifice and feast reflected the killing and eating of the father, assuaged "the burning sense of guilt," and brought about a kind of reconciliation or agreement by which the father-totem granted all wishes of his sons in return for their pledge to honor his life (2, p. 237). "All later religions prove to be . . . reactions aiming at the same great event with which culture begins and which ever since has not let mankind come to rest" (2, p. 239).

VI

Kroeber (3) has offered a criticism of this elaborate theory; a partial summary is given here. In the first place, the Darwin-Atkinson hypothesis of the "Cyclopean family" is almost wholly conjectural, if not disproved.³ Robertson Smith's thesis that blood sacrifices are central in ancient cults holds only or chiefly for Mediterranean cultures of the last two millennia B. C. and for cultures subsequently influenced by them. It is at best problematical whether blood sacrifice originated in a totemic observance.

Turning to immanent criticism of the Freudian theory proper: It is only a guess that the sons would kill, much less devour, the father. The fact that a child sometimes (*in how many cases?*) displaces his father-hatred upon an animal affords no proof that the sons did so. Even if such displacement took place, would they retain enough of the original hate-impulse to slay the father, and if so, would not the slaying dissipate the displacements? And, granting the remote possibility of the sons' remorse and consequent

³The hypothesis has been out of favor, but recent studies of anthropoids by biologists and mammalogists indicate possible confirmation. On this point see V. F. Calverton (1).

resolve no longer to kill the father-displacement-totem, it seems dubious, to say the least, that this resolve would be powerful and enduring enough permanently to suppress the gratification of the sexual impulses possible after the killing of the father. If the brothers allowed strangers (perhaps expelled by their own jealous fathers) access to the women they themselves renounced, thus perhaps bringing into existence matrilineal or matriarchal institutions, what would be left for the brothers, unless they could be content with life-long celibacy or homosexuality, other than individual attachments to other sibs—attachments which would mean disintegration of the very solidarity they are pictured as so anxious to preserve even at the cost of denying their most deep-rooted physiological urges?

Moreover, it has by no means been proved that exogamy and totem abstinence are the two fundamental prohibitions of totemism—quite the contrary. Freud cites Goldenweiser's study of the subject, but neither accepts nor refutes the latter's conclusion that these two features are not generally basic in the totemic complex. Once more, these two totemic taboos cannot be shown to be the oldest of all taboos, and no presentation of the nature and sequence of the displacements and distortions said to produce all other taboos from these is given. Finally, the persistence into modern society and religion of "this first great event with which culture began" is an unexplained process.

In spite of this devastating criticism, Kroeber approves of several points in the theory, some of them closely linked with the manifestation of ambivalence. The correspondence between taboo customs and "compulsion neuroses" as developed by Freud, and the parallelism between the two aspects of taboo and the ambivalence of emotions under an accepted prohibition, Kroeber holds to be indubitable. "The strange combination of mourning for the dead with the fear of them and taboos against them is certainly illumined if not explained by this theory of ambivalence. It is even possible to extend Freud's point of view. Where the taboo on the name of the dead [to be discussed later] is in force we find not only the fear that utterance will recall the soul to the hurt of the living, but also actual shock at the utterance as a slight or manifestation of hostility to the dead" (3).

VII

Malinowski, in his book, *Sex and Repression in Savage Society*, sets forth a different type of criticism resulting from an effort to test the Oedipus-complex hypothesis in a matrilineal society. Like Kroeber, he finds the Freudian theory thoroughly unsatisfactory as a detailed explanation although some points may be substantiated. Malinowski's misapprehension (in the first part of his book) of the meaning of the term "complex" as used by Freud by no means adds to the force of his criticism, to be sure, but nevertheless its main outlines are valid.⁴

Among the Trobriand Islanders the physical facts of fatherhood are apparently unknown; kinship in the social sense exists only through the mother; the father is a beloved, benevolent friend but nothing more. Authority is vested in the mother's brother, who is the chief breadwinner. Marriage is patrilocal; the home, the traditional center of local patriotism, the family possessions, and the pride of ancestry are not in the place where the children live. The strong brother-sister taboo forbids any common interest in sex matters, so that brothers and sisters are separated at an early age (6, pp. 8 ff.).

Without going into details of social psychology as affecting the Trobriand family, it may be said that Malinowski finds no trace of the Oedipus complex discovered by Freud in a patrilinear, patriarchal European culture and assumed by him to be universal. Malinowski's main conclusion is that "a matrilineal family system arises, for unknown social and economic reasons, and then the repressed nuclear complex consists of brother and sister attraction, with nephew and uncle hatred; when this system is replaced by a patrilineal one, the nuclear complex becomes the familiar Oedipus one" (6, p. 142). Ernest Jones, following Freud, sees in this merely a repression of the Oedipus complex: "the forbidden and unconsciously loved sister is only a substitute for the mother, as the uncle plainly is for the father" (6, p. 143), although as Jones admits, there is no trace of the typical Oedipus complex to be found in Trobriand folk-lore, dreams or visions (6, p. 144). Malinowski further criticizes the Freudian theory along the general lines of Kroeber's approach. His most telling point is that the parricide which Freud says was

⁴For a discussion of *Sex and Repression in Savage Society* from the standpoint of method, see Harold D. Lasswell (4).

the starting point in "culture" was, according to Freud's own words, "perhaps" made possible because of "some advance in culture, like the use of a new weapon." Thus Freud, as Malinowski says, "equips his pre-cultural animals with a substantial store of cultural goods and implements," and "no material culture is imaginable without the concomitant existence of organization, morality, and religion" (6, pp. 152-153). Freud further asserts the presence of a "mass psyche" endowed with an almost unlimited biological memory permitting the continuance of emotional processes arising in generations of ill-treated sons in all succeeding generations which by the very removal of the primordial "Cyclopean" father necessarily escape such treatment. Malinowski properly points out that the "collective sensorium" is an unnecessary and thoroughly discredited concept.

VIII

Mention should here be made of a further ethnological criticism of *Totem and Taboo* by Margaret Mead (7), who discusses the hypotheses that, granting ambivalence toward the dead, the dominant aspect of this ambivalence is a function of the civilization in which an individual lives, and that it is possible to find a cultural solution of these conflicting attitudes which may obviate the need of suppressing either one.

It is highly significant that Kroeber and Malinowski reject the primal horde theory and the exclusive importance of the Oedipus complex, while each accepts other important elements of the Freudian theory. The former accepts the correspondence of taboo practices and compulsion neuroses, and the importance of ambivalence; the latter, the fundamental place of a nuclear complex connected with the form of the family. Similarly significant is Mead's assertion that ambivalence toward the dead, in some instances at least, may be primarily a cultural rather than a biological function.

IX

We have now defined our terms and have given an abstract and several important critiques of Freud's theory. The field is clear for a systematic statement of a hypothesis which, although similarly conjectural, seems to avoid the objections just raised. Above all, it is not offered as a comprehensive explanation of all totemic or taboo phenomena, nor does it make totemism and exogamy interdependent.

First, certain negative elements or qualifications must be stated. In each preliterate group only one or two persons, a few at most, were what we should call neurotic or psychotic; this is in line with what was previously stated with regard to ambivalence. It follows logically that the Oedipus or other basic family complex was and is not a factor exercising a marked influence on behavior in more than these very few persons: the complex is a symptom of neurosis, or the neurosis may be caused by the complex, the two being interdependent. The vast majority are undisturbed by any such emotional conflict, although the difference between the "normal" person and the neurotic is one of degree, not of kind. These negative aspects are all points of difference from the theory of Freud.

We have suggested in the earlier discussion of ambivalence how excesses of grief or mourning behavior may result from emotional ambivalence and how the mechanism of projection may produce ghost-fear. From these concepts of ambivalence and projection, carried a little further, we may derive the concept of compulsion or anxiety neurosis as an explanation—tentative, provisional, and partial—of taboos of the dead and of the property and name of the dead. Here we follow Freud with the important qualifications previously noted.

Compulsion neurosis (or similar psychosis), as the name indicates, is a nervous condition in which a person is compelled to carry out apparently unmotivated behavior: to arrange his personal effects in a particular pattern before going to bed, for instance, regardless of time, place or social convention; or under all circumstances to avoid touching an object such as a door-knob. In nearly all cases the sufferer will offer a highly "rational" explanation.

The points of correspondence most clearly manifest between taboo customs and compulsion neurosis are as follows: (1) in the lack of motivation for the commandments, taboos or compulsive actions, which appear "senseless" to the outside observer; (2) in their enforcement through an inner need of the neurotic individual; (3) in their capacity for displacement or substitution (by re-conditioned response or "transfer") and in the danger of "contagion" from what is prohibited; and (4) in the causation of ceremonial actions and commandments which emanate from the forbidden (2, p. 48). From this it would appear that the basis of taboo, as stated in our earlier discussion of the term, is a forbidden action for which there exists a strong inclination in the unconscious.

A hypothetical case will help to make these points clear. Let us assume that as a result of experiences in early life a boy comes to entertain a strong death-wish against his father based on jealousy for the affections of the mother, a wish he cannot gratify because of the father's superior power or the already existing tribal taboos. If the conflict, because of its unusual intensity or the child's unstable nervous organization, causes neurotic symptoms, he may become as a consequence an amateur or professional practitioner of magic or religion—medicine-man, shaman, priest, etc. (The term "priest," while not wholly accurate, will be used hereafter.) When the father dies the death-wish of the son is satisfied so far as outward circumstances are concerned. But here emotional ambivalence introduces conflicting elements: along with gratification at the realization of the death-wish there is also a certain distress, since (according to the hypothesis of ambivalence) a certain amount of tender feeling has persisted beside the hate or the death-wish, and this love causes distress or sorrow.

Now, the hate elements in the unconscious of this priest might cause corpse-fear and ghost-fear, through the mechanisms of projection previously spoken of, and these elements alone might cause *taboo through fear*. The incest-wish (to use Freud's phrase) would not be gratified and the property of the father would not be appropriated by the priest because of his fear of the ghost, which is really the projection, made possible by animistic belief in a spirit-double, of his own hatred.

There is, however, another and perhaps more powerful reason for taboo of the property and the widow of the dead which might operate in such a case. The tender feeling which had always persisted to some extent by the side of the priest's hate for his father might gain the upper hand as soon as this hate was satisfied and cancelled out by his death; and it might then take the form of remorse and sense of guilt, with resulting fear of the ghost, to whom revenge would be imputed. The incest which the superior might of the father had formerly prevented the son himself might now prohibit, unconsciously or consciously, resulting in the psychic situation of "subsequent obedience" of which psychiatrists make so much. His previously entertained death-wish amounted to a virtual killing of the father, so far as its unconscious effects are concerned, and he might now renounce the "fruits" of his "deed" by denying himself access to the wife-mother. This "good resolution" might

be extremely feeble, however, if sex desire were strong enough, and the continual temptation to which the shaman was thus exposed might be further strengthened by the example of others who were not under the same compulsion to refrain from sexual intercourse with his father's widow.

At this point we have a reason for the introduction of taboo of the wife of a dead man by direct prohibition as well as by example. The priest-son who is exposed to temptation in the manner just mentioned may as a consequence prohibit such behavior by others. This finds an analogy in the behavior of many neurotics, who declare various modes of behavior to be "impossible"—who will not only restrain themselves from various entirely natural acts, but who, if they had the power, would restrain all others from the same acts. The way in which the priest would restrain his fellow-tribesmen from intercourse with his father's widow would depend on the particular rationalization employed by him—he might say that she was likely to kill the men with whom she came in contact, or that the ghost would revenge itself on those who violated its "property," or that contact with a tabooed person or thing, by virtue of a power of "contagion" residing in them, would cause the individuals indulging in such contact to become themselves taboo or unclean. Such injunctions and prohibitions, by virtue of the power wielded by the neurotic priest, would be obeyed to the letter and be handed down through centuries as part of the culture-pattern of the tribe. Other neurotic priests might suffer in similar fashion, and add various elements of taboo to the culture-pattern, until, in the case of tribes with relatively complete cultural continuity, life might become a network of more or less rationalized taboos and restrictions. As a matter of fact, however such a situation comes to be, life in most preliterate groups *is* such a network.

In this connection we may also consider taboos on the names of the dead. As Kroeber points out, "Where the taboo on the name of the dead is in force we find not only the fear that the utterance will recall the soul to the hurt of the living, but also actual shock at the utterance of a slight or manifestation of hostility to the dead. It is a fair question whether this shock may not be construed as a reaction from the unconscious hate carried toward the dead during their life, *as if speaking of them were an admission of satisfaction at their going*. The shock is certainly greatest where affection was deepest; persons who were indifferent are mentioned without

emotional reluctance if circumstances permit, whereas enemies toward whom hate was *avowed* may have the utterance of their names gloated over" (3, pp. 52-53).

Such taboos may go to the length of changing the names of surviving near relatives, as in some California and Pacific Coast and some New Guinea tribes. A related observance is the injunction to say nothing but good of the dead, which operates in both pre-literate and modern social groups.

X

Now, one possible way, among many others, by which taboo and totemism might be linked together, is through the addition *at different times* of taboo and totemic elements to the culture-complex. They need not be interdependent, but because of similar psychological origins may seem so.

For instance, another priest might have an Oedipus complex similar to that just described, but in addition the psychological phenomenon known as unconscious displacement may have taken place. That is, the hatred repressed in the unconscious of the child may have found partial release through being vented on a symbolic father-substitute. This father-substitute might be another man, a woman, an animal, or in fact any object which would serve as a symbol for certain characteristics of the father, and upon which hatred could be projected. The hatred rooted in the death-wish, which was really entertained against the father, might be gratified by killing or torturing the animal, let us say, hit upon as a symbol for certain characteristics identified with him. Such displacement-behavior actually takes place in a few neurotic children whose cases have been studied, so this possibility is not mere conjecture. The animal may be killed and eaten, in which symbolic act the death-wish against the father finds its final consummation, so that when the father dies the everyday act of killing and eating the animal is equivalent to an explicit avowal of hatred. This, as we have already seen, is seldom done because of the ambivalence present; in fact the sense of remorse and guilt arising upon gratification of the death-wish frequently leads to a direct prohibition or taboo of the animal formerly identified with the father, and this really is an attempt to "unwish" or "call back" the earlier death-wish.

Such prohibition of an animal identified with the father may conceivably pass over into one of the fundamental marks of totem-

ism: an animal is regarded as the father of the tribe, and this animal may not be killed or eaten. A member of the Kangaroo totem "knows," as a result of his early training, that a Great Kangaroo was the ancient father of his totem, and "knows" as a consequence that kangaroos are forbidden to him. This belief may have originated with the neurosis of some primitive priest who enforced his own inner compulsion upon the rest of the tribe, in the manner already dwelt upon.

The taboo of the dead man's widow we have assumed to be already in force as a result of the neurosis of some earlier priest; so that when this neurotic totemism is added to the culture-complex, there are two elements in it, taboo and totemism, which have had an independent origin with similar psychological mechanisms operating in each case. What wonder if they appear interdependent and inseparable? (Let us repeat, however, that no hypothesis here advanced is offered as more than a *partial* explanation of very complex phenomena.)

Further, when both elements are included in the culture-complex, there is room for possible extension of their influence by interaction. The fear of incest with the mother may grow to fear of sexual intercourse with anyone who could possibly stand in blood or social relationship to the father. It will be noted that a monogamic family has been assumed up to this point. Now, if any one of the more complex forms of marriage (group marriage, for example) prevails, nearly everyone in the group may be in some distant way connected with the father (or with the person barring access to the mother—whether or not he is the actual father is immaterial). It is at least probable that some neurotic priest may have added taboo of all women within a certain degree of consanguinity to the culture complex, and when the totemistic belief described above is added to this, we have a situation very nearly like that found, for example, in certain Australian tribes. As various exogamous groups became connected by intermarriage, a system of mutual relationships would slowly grow up, and conceivably might become as intricate, through later rationalization upon what was at first culturally compulsive or imitative, as the phratry, sub-phratry, and totem structure of the groups mentioned.

There is another element of totemism which might come to form an integral part of the slowly-built-up system described, and that is the ritual sacrifice of the totem animal and eating of a part of it

as an expiation or atonement for some "sin" or "crime" against the Great Totem-Father of the tribe. This sinful "act" symbolizes the death-wish and its correlate incest-wish, and need never have taken place in actual fact. By becoming one with the totem through the mystic process of eating a part of it, the worshipper is afforded a sense of identification with it, a consequent "holiness," and a release from the sense of guilt. This whole process is too complicated and the various controversial issues involved too far-reaching, however, to be further discussed within the limits of this article.

XI

In presenting this alternative theory we have used throughout a hypothetical instance involving the familiar Oedipus complex of a patrilineal society. In a matrilineal society, if we may accept the observations of Malinowski, the psychological mechanism would be the same, with substitution of the maternal uncle for the father, and of the sister for the mother.

No far-reaching claims are made for this hypothesis of the origins and transmission of taboo and totemism. It is not intended to explain all phenomena of the sort, but merely offers a partial and tentative explanation. Epitomized, the theory is as follows: neurotic or psychotic individuals, who (perhaps because of their abnormal condition) held the rank and authority of shamans, medicine-men, priests, etc., prescribed for themselves and others (who otherwise would cause too much temptation) certain taboo restrictions which took the form of compulsion neurosis in the priest and in the others of mere external commands enforced by the priest's supposed power over the extranatural or supernatural world. These taboos became an integral part of the group's tradition, and other neurotics later added in the same manner the "displacements," transference or reconditioning of totem-belief, with subsequent extension of the influence of both totem and taboo to the group life of the whole tribe, ending in rites of piacular sacrifice to the great Father-Totem.

The merits of this hypothesis are these: (1) most important, it avoids Freud's unprovable assumption that preliterate ("primitives") were neurotic to the extent that all were afflicted by compulsion neuroses corresponding to taboo customs; (2) it does not pyramid hypotheses in an attempt to make taboo and totemism interdependent and of contemporaneous origin; (3) it is flexible enough to account for the presence or absence of one or another of

the various elements in the totem-taboo combination. There may be taboo without exogamy, totemism without taboo or exogamy, taboo without sin-offerings or sacrifice, totemism without worship of the totem, or a grand *mélange* of all these factors in one culture-complex. Freud's hypothesis stands or falls on the unfailing connection of totemism with exogamy, a connection at best flimsy and more probably disproved.

REFERENCES

1. CALVERTON, V. F. The compulsive basis of social thought: as illustrated by the varying doctrines as to the origins of marriage and the family. *Amer. J. Sociol.*, 1931, **36**, 698-720.
2. FREUD, S. Totem and taboo. (Trans. by A. A. Brill.) New York: Moffatt, Yard, 1918. Pp. 297.
3. KROEBER, A. L. Review of Freud's *Totem and Taboo*. *Amer. Anthropol.*, 1920, **22**, 48-55.
4. LASSWELL, H. D. A hypothesis rooted in the preconceptions of a single civilization tested by Malinowski. In *Methods in Social Science*, ed. by S. A. Rice. Chicago: Univ. Chicago Press, 1931. Pp. 480-488.
5. LOWIE, R. H. Primitive religion. New York: Boni & Liveright, 1924. Pp. 346.
6. MALINOWSKI, B. Sex and repression in savage society. New York: Harcourt, Brace, 1927. Pp. 299.
7. MEAD, M. An ethnologist's footnote to *Totem and Taboo*. *Psychoanal. Rev.*, 1930, **17**, 297-304.

Smith College
Northampton, Massachusetts

University of Pennsylvania
Philadelphia, Pennsylvania

QUELQUES ASPECTS DU TABOU ET DU TOTÉMISME

(Résumé)

L'ambivalence émotive, par les mécanismes de la projection et du déplacement, donne une explication partielle du tabou et du totémisme mais non de la manière expliquée par Freud. La supposition par celui-ci d'une "horde primale" et l'universalité du complexe d'Oedipe, menant à l'assassination et au manger du père comme l'acte basique menant au tabou et au totémisme, s'est montrée non substantiée, pour des raisons théoriques par Kroeber et par les observations de Malinowski et de Mead sur les sociétés matrilineaires. Nos auteurs offrent ici une hypothèse libre de ces suppositions. Pour établir une observance du tabou ou du festent l'ambivalence émotive par la projection ou le déplacement, surtout dans la position de prêtre. Ainsi des tabous sexuels sur la veuve du père, des tabous sur le nom des morts, ou (par le mécanisme du déplacement) le symbole totémique et l'abstention de les manger, peuvent être établis pour le groupe entier par des membres psychopathiques. Des éléments

séparés de temps en temps par des individus peuvent produire précisément un tel réseau de tabou et de totémisme apparemment dépendants l'un de l'autre, comme on le trouve en effet dans la société primitive. On ne dit pas que cette hypothèse est complète, laquelle offre une explication partielle des phénomènes très compliqués.

BECKER ET BRUNER

EINIGE BETRACHTUNGEN ÜBER TABU TOTEMISMUS

(Referat)

Die affektive Ambivalence (emotional ambivalence) bietet uns durch die Mechanismen der Projektion (projection) und der affektiven Verschiebung (displacement) eine partielle Erklärung des Tabu und des Totemismus dar, aber nicht auf die von Freud entfaltete Weise. Freud's Annahme einer Urhorde (primal horde) und eines allgemeinen Oedipus-komplexes der zur Tötung und Verzehrung des Vaters führt, ist als grundlegende, zum Tabu und zum Totemismus führende Tatsache, aus theoretischen Gründen von Kroeber und durch die Beobachtungen von Malinowski und von Mead an matriarchalischen Gesellschaften als unbewiesen dargestellt worden. Der gegenwärtige Verfasser bietet eine Hypothese an, die von diesen Annahmen frei ist. Ein Paar Psychopathen, die affektive Ambivalence durch Projektion oder affektive Verschiebung offenbaren, besonders im Priesterstand, genügen, um in der ganzen Gruppe ein Tabu oder ein totemisches Ritual einzuführen. Auf diese Weise können Geschlechtstabu (sex taboos) gegen die Witwe des Vaters, Tabu gegen den Namen eines Toten oder (durch den Mechanismus der Verschiebung), das totemische Symbol und die Abhaltung von der Verzehrung desselben für die ganze Gesellschaft durch pathologische Mitglieder aufgerichtet werden. Die Hinzufügung einzelner Bestandteile durch Individuen von Zeit zu Zeit kann gerade ein solches Netzwerk von scheinbar gegenseitig abhängigem Tabu und Totemismus zu Stande bringen wie man es tatsächlich in der primitiven Gesellschaft findet. Es wird nicht behauptet, dass diese Hypothese vollständig sei; sie bietet nur eine partielle Erklärung für sehr verwickelte Erscheinungen.

BECKER UND BRUNER

SHORT ARTICLES AND NOTES

CONCERNING THE THURSTONE "PERSONALITY SCHEDULE"— ERRATUM

O. L. HARVEY

Footnote 1, acknowledgment to Mr. Carl A. Fehr, which was published in connection with the article entitled "The Scientific Study of Human Sexual Behavior," this journal, page 161, Vol. III, No. 2, May, 1932, should have appeared in connection with the short article entitled "Concerning the Thurstone 'Personality Schedule,'" page 240 of the same issue.

THE MICHIGAN NON-VERBAL SERIES

A NON-VERBAL TEST SERIES FOR MEASURING HAND AND EYE COORDINATION,
OBSERVATION, AND COMPARISON, FROM FOUR YEARS TO SUPERIOR ADULT

EDWARD B. GREENE¹

The attempt to compare and measure human performance inevitably involves four fundamental considerations: (*A*) the equivalence of measuring units, (*B*) the equivalence of individual efficiency, (*C*) the equivalence of sample populations, and (*D*) the equivalence of zeros in the scales used. At present there seems to be no direct means of equating any of these, but certain methods and criteria may be employed to give more equivalence than otherwise. Thus, for equating measuring units, one must strive to make the processes uniform with regard to complexity and range of information and the physiological bases. For equating individual or group efficiency one must try to secure equal motivation, equal maturation, equal practice. For equating populations there are both statistical and social criteria, such as the shape and size of the frequency polygon, the cultural environments and the languages used. For equating zeros, or the beginning of development of the factors contributing to success in a certain performance, one must measure as near to the zero point as possible, then calculate the unmeasured portion of the scale—much as the physicist calculates the absolute zero in temperature.

The Michigan Non-Verbal Series² has been developed to meet these considerations particularly in the measurement of observation and com-

¹This work was done in part while a Fellow of the Social Science Research Council, 1930-1931.

²Published by Edward Bros. Co., Ann Arbor, Mich., 1931.

parison of printed objects. It is designed to reduce to a minimum the effect of differences in information and to deal particularly with the comparison of simple space patterns.

A. In order to secure some equivalence of test units the following specifications were followed.

1. *Nature of Tasks.* The tasks should be (1) of such a sort as to demand careful observation and comparison, (2) of such a sort as to avoid differences due to cultural backgrounds, and (3) interesting enough to call out the subject's best efforts.

2. *Non-verbal Test Objects.* Non-verbal material is specified in order (1) to allow the series to extend from small children to adults in all language groups, (2) to show reasoning ability independent of verbal and number information, and (3) to allow for increases in complexity without bringing in new information.

3. *Pencil-and-Paper Situation.* This was adopted (1) because of the ease in administering and scoring, (2) because a very large proportion of persons to be measured would have some skill in using a pencil, and (3) because this skill as it applies to the test situation can be quickly and reliably measured.

4. *Four Levels of Complexity Recorded Separately.* This allows (1) an analysis of one's responses to similar tests of varying difficulty, (2) an analysis of the normality of dispersions of various age groups at each level of difficulty, and (3) the analysis of speed vs. power in individual performance.

5. *Separate Evaluation of Physiological Adjustments Which Are Basic to the Test Situation.* This will give (1) independent measures of vision and motor coordination which will show to what extent the measures of observation and comparison are dependent on these, and (2) a reliable indication of handedness.

7. *Visibility.* (1) The test objects must be so large and clear that they can be easily seen at twelve inches by a person with 20/30 vision, uncorrected, with either eye, under an illumination of three-foot candles. These conditions are commonly found in classrooms. (2) The print should be of light green or blue ink on a dull-finish white paper, since this combination gives few after images, but good contrast.

B. In order to secure equivalence of individual efficiency these rules were followed:

1. *Repetitions with Alternative Forms until a High Level of Efficiency Is Reached.* This is included because it has been found that the first and second trials in this series are only fairly good indications of the fourth trial. Large preliminary differences in motivation and training were found which tend to disappear with practice. The correlations of first and fourth trials range between .45 and .75; of third and fourth trials, between .76 and .94 for age groups from 4 to 29 years. It is therefore

recommended that four trials of the series be used at intervals of about one day.

2. *Work Periods Sufficient for a Good Sample of Performance without Much Fatigue.* This specification serves to limit the test to the periods generally available and to keep motivation fairly high. For certain purposes it would be well to supplement this with endurance tests.

C. In order to secure equivalent samples of a white population at various ages the aid of Dr. H. J. Baker of the Psychological Clinic of the Detroit Public Schools was secured. At his suggestion the Keating School was chosen for the elementary grades because the records showed that this school was one of the most representative of the white Detroit population in range and distribution of school ability, racial and language backgrounds, and economic status. For the older age groups fairly adequate samples of persons in the senior high schools, College of the City of Detroit, evening grade schools for adult and the Ford Trade School, and the Illinois State Penitentiary at Joliet were secured. In this manner approximately 150 persons were measured in each age group from 5 to 20 years.

D. In order to secure evaluations of zeros which represent the beginnings of development, two methods have been devised by Thurstone (1); one is the extrapolation of the known curve of age group averages and the other is the calculation of a point on the scale where the dispersion of a normally selected group becomes zero in the trait measured, on the assumption that the factors which operate in the known segment will also operate in the unknown segment. These methods gave nearly the same results with the material of this series based on the fourth trials of about 2000 white persons. The absolute zeros are incorporated in the published norms.

II. DESCRIPTION OF THE TESTS

In order to evaluate the visual and motor factors three tests have been devised: a test of visual acuity and astigmatism, a test for accuracy of aiming, and a test for speed of movement. Accuracy of fine coordination and speed have been measured independently because they vary independently to a large extent, and because they are thought to have some vocational predictive value.

The visual acuity charts are of two sizes, one for 6 meters or 20 feet, and the other for 30 cm. or 12 inches. Each consists of eleven rows of circles crossed by grids, in which the grid lines subtend an angle of one minute at the distance indicated. Each row has six circles and grids of the same size which are placed in the six directions of the clock hours in random order. The task is to show by pointing or naming clock numbers the directions of the grid lines as far down the chart as possible. The

charts give a good indication of astigmatism for both near and far distances. They may be applied to the average four-year-old.

The aiming sheet consists of rows of small circles into each of which the subject must put a single dot. The circles are 1.5 mm. in diameter. Two periods of 60 seconds are allowed for each hand. The number attempted and correctly marked is recorded for each hand separately. Accuracy is stressed in the directions.

The speed-of-movement sheet is divided into six squares. The subject taps as fast as possible for ten seconds with the right hand in each of the three right-hand squares and with the left hand in the three left squares. The number of dots are counted by drawing circles around groups of five dots until all are included. No reliable or constant differences have been found among trained subjects under controlled conditions at various ages from four to twenty years between tapping with a stylus on a brass plate, and with pencil on paper, for ten-second periods. Movement has been controlled only to the extent of having both elbows supported on a desk which approximates the height of the costal arch apex.

In the observation and comparison series, two types of test objects have been used: small round faces and pencil mazes. These were chosen because they conformed to the specifications listed above for equivalence of units, and because they require different methods of solution. The maze is a sort of pursuit problem, while the feature discrimination depends more on a methodical proposal and checking of hypotheses.

The faces are arranged in rows. The task is to find one unique feature in each row, and indicate it by drawing a small circle around it. Each face has four features, hair, eyebrows, eyes, and mouth, which vary in a random order in size, direction, curvature, and, in the case of hair, number. The easiest page calls for the comparison of two faces, the next, three faces, the next five, and the hardest, seven faces in a row. The number of rows attempted and correct during fairly liberal time limits are recorded for each page.

The pencil mazes were constructed after considerable experimental work to give the maximum difficulty of solution with the minimum of chance and motor factors. They have but one true path and they have proven to be equally difficult of solution in either direction. This allows for left-handed variants and makes it impossible to solve faster by back tracing. Four levels of difficulty have been prepared on separate pages ranging from very easy to very difficult. The amount of true path and number of blinds entered are recorded for each level.

III. ORDER OF TESTS

The aiming series has been placed first because it requires sharp pencils and a minimum of fatigue. The tapping test is also placed on the first page to allow a direct comparison of motor tasks. The feature discrimi-

nation series comes next, with four separate pages each of which is more difficult than the last. The maze series are placed at the end because they seem to have more intrinsic interest than the others for all ages considered. Four complete repetitions of the entire series are found in each booklet—a total of 36 pages. The visual acuity charts are furnished separately.³

IV. ADMINISTRATION OF TESTS

The total working time is 26 minutes for adults and 16 minutes for children below the nine-year level. The total administration time for groups should allow about 35 minutes. When pantomime directions are used, as for deaf, or persons speaking different languages, approximately 60 minutes should be allowed. Individuals below the sixth-year average should be tested individually. When individuals are tested separately these time limits are generally reduced about 10 per cent. On repetitions of alternative forms no samples are needed, therefore the administration time is only slightly longer than the working time.

The group administration calls for good voice as well as careful timing.

It can be mastered in two or three practice periods by the average college sophomore. Each booklet includes directions for administering and scoring the series. The test for visual acuity must always be an individual test administered by a carefully trained person.

V. TEST NORMS

This series has been applied to approximately one hundred and fifty persons at each age from 4 to 20 years, selected so as to give a good random sample of the white population in the Detroit area, excluding institutional populations. The distribution of scores of each twelve-month group are given on a graphic profile sheet. The individual's records on the various parts of the series are recorded separately on this sheet and interpreted in terms of mental age, IQ, centiles, or standard deviations (*S.D.*) of his age group. The *S.D.* score is recommended because it has considerable current usage, greater equivalence of units than the centile or age means, greater range than the age means, and more statistical significance than the IQ as customarily used.

VI. SUMMARY

The Michigan Non-Verbal Series has been developed to measure vision, hand and eye coordination, and observation and comparison of simple objects. It incorporates the following:

1. Uniformity of tasks to give equivalent measuring units. The emphasis is placed on observation and comparison rather than information.
2. Wide range of complexity without changes in objects or directions. Each level of complexity measured separately.

³E. B. Meyrowitz Co., New York, N. Y., 1931.

3. Four repetitions of equivalent forms to secure more equivalence of individual efficiency, and records of improvement and variability.

4. The selection of fairly adequate samples of the Detroit white population from 5 to 20 years of age, and norms for each age group on each repetition of the series.

5. The calculation of absolute zeros in each scale and their use in a graphic profile sheet for individual scores.

REFERENCE

1. THURSTONE, L. L. The absolute zero in intelligence measurement. *Psychol. Rev.*, 1928, **35**, 175-197.

University of Michigan
Ann Arbor, Michigan

A METHOD OF CORRECTING FOR GUESSING IN TRUE-FALSE TESTS AND EMPIRICAL EVIDENCE IN SUPPORT OF IT

KATE HEVNER

One of the most important objections to the true-false type of test is the large unearned increment which the subject acquires for himself by pure chance. Various methods of scoring these tests in order to counteract this effect of chance have been advocated. Experimental evidence has been brought forward to show that at least one of these methods, the Right-Minus-Wrong scoring formula, has not proved satisfactory in improving the reliability or the validity of the tests (4). The present study offers some empirical evidence in support of a method of scoring which gives weight to the items in the test according to the degree of confidence which the subject feels in his responses.

The test material used in this study comes from two tests in the field of aesthetics, the Meier-Seashore Art Judgment Tests (3), an hour test composed of 125 items and the author's test for appreciation of music (2), an hour test, of 48 items. The reliability coefficients for both tests are somewhat below the accepted standards for psychological tests, and since, for the music tests, the nature of the material necessitates the true-false form, an empirical study of four different scoring formulae was made in the hope that a correction for the effect of chance might materially increase the reliabilities.

In the printed directions to the subjects for these tests, a new factor, the subject's confidence in his judgment, was added. The directions were quite explicit, and a sample marking was provided as follows:

You will hear some piano compositions in groups of two short pieces which are very much alike. Listen carefully and decide which of the two you prefer, that is, which of the two sound most pleasing, most appropriate, most musical to you. If you prefer the first one, put a circle around "1" in the space follow-

ing the number of the piece. If you found the second one most pleasing put a circle around "2." In other words, encircle the number of the one you prefer.

In some cases you may find it very easy, and in some cases very difficult to decide, and we are interested to know the degree of confidence you feel in your judgment.

If you feel that you would always make the same choice whenever you heard the two pieces, put a circle around the A. (Very sure.)

If you feel that you might possibly change your mind if you could hear the two pieces played again put a circle around the letter B. (Fairly sure.)

If you feel that there is really no choice between the two, and you chose one or the other only because you had to mark something, put a circle around C. (Not at all sure.)

Here is an example. If you were *very sure* that you preferred the first one for number 37, your record would look like this:

Composition
Number
37

Preference
① 2

Degree of
Confidence
Ⓐ B C

The scoring formulae applied to the tests were the following:

1. *Number Right*: The sum of the correct choices.
2. *Right-minus-Wrong*: The subject was penalized for all his errors by having them subtracted from his number right.
3. *Weighted Right Answers*: The right answers were weighted according to the degree of confidence which the subject felt toward his response. A response graded as A counted as 3 points in the score, B counted 2 points, and C one point.
4. *Weighted Right Minus Weighted Wrong*: The subject was not only rewarded for a right response graded A, but penalized two points for a wrong response graded A, penalized one point for a wrong response graded B, but not penalized at all for having "guessed" wrong as indicated by a C grade of confidence.

The first study was made on the music appreciation test, with one hundred freshmen and sophomores in the University of Minnesota, who took the test in March, 1931. The subjects were students in the elementary psychology course or in the elementary course in music theory. Before taking the test, these subjects filled out a self-rating scale for music talent, and a more objective scale for music training (1, p. 31). These scales served as the criteria for the validity of the test. The results from this first experiment are given in Table 1. Since these coefficients of reliability were calculated from the odd-even scores, the corrected coefficients for a test of twice as many items, according to the Spearman-Brown formula, are included. The advantage for both reliability and validity goes to the weighted right answers.

TABLE 1
RELIABILITY AND VALIDITY COEFFICIENTS AND THEIR PROBABLE ERRORS FOR 102
SUBJECTS OF THE HEVNER TEST FOR APPRECIATION OF MUSIC

Scoring formula	Odd-even	Estimated reliability (Spearman- Brown)	Talent	Training
Number right	.502±.05	.67	.41±.04	.50±.04
Right-minus-wrong	.49 ±.05	.65	.45±.05	.67±.04
Right weighted	.71 ±.03	.83	.53±.05	.70±.04
Right weighted minus wrong weighted	.62 ±.04	.76	.49±.05	.67±.04

Both the music appreciation test and the Meier-Seashore Art Judgment Test were given to the 110 students registered in the elementary laboratory course in the spring quarter of 1931. These students represent an extremely homogeneous group in that they are practically all sophomores from the Arts College, with no representatives from the professional schools, and that there is very little deviation from the median age of 20 years. This homogeneity means that the range of the scores is more restricted than in the first study, and that the coefficients are all somewhat lower. The two rating scales are again used as the criteria for validity, although, in the absence of such a scale for the art test, only the reliability coefficients are reported.

The advantage goes in every case to the weighted right answers, and the gains are quite material especially in the case of the Meier-Seashore test when the reliability is increased from .50 to .66. It is possible that some slight variations in the weighting might result in an even larger gain in reliability, or empirical data may show that an equal gain might be effected if the subject used only two different degrees of confidence.

TABLE 2
RELIABILITY AND VALIDITY COEFFICIENTS AND THEIR PROBABLE ERRORS FOR 100
SUBJECTS ON THE HEVNER TEST FOR APPRECIATION OF MUSIC AND THE
MEIER-SEASHORE ART JUDGMENT TEST

Scoring formula	Odd-even music	Estimated reliability music	Talent	Training	Odd- even Art	Estimated reliability Art
Number right	.39±.06	.56	.33±.06	.48±.05	.50±.05	.66
Right minus- wrong	.42±.06	.60	.34±.06	.51±.05	.51±.05	.67
Right weighted	.54±.05	.70	.44±.06	.57±.05	.66±.04	.80
Right weighted minus wrong weighted	.39±.06	.56	.32±.06	.46±.05	.56±.02	.71

Further studies with these variables are indicated, not only with these tests, but with any other material in the true-false form. The device is certainly not limited to tests in the field of aesthetics but applies directly to any test where a time limit is not rigidly imposed, and very possibly will prove of value to speed tests if a small allowance of time be made in addition to the standard testing time.

Informal observation among the subjects indicates that the opportunity to express a degree of confidence is a welcome addition to the test, especially when the feeling is one of insecurity. The subjects in these tests were not told what the method of scoring would be since the experimenter did not know in advance which formula would have the advantage. It is undoubtedly true that if the method became generally known dishonest subjects could artificially raise their scores.

REFERENCES

1. FOSTER, W. S. & TINKER, M. A. Experimental psychology. Notes for instructors. New York: Holt, 1929.
2. HEYNER, K. Tests for aesthetic appreciation in the field of music. *J. Appl. Psychol.*, 1930, 14, 470-477.
3. MEIER, N. C., & SEASHORE, C. E. Art Judgment Test. Iowa City, Iowa: Univ. Iowa, Bur. Educ., 1929. Pp. 125.
4. PATERSON, D. G., & LANGLEY, T. A. Empirical data on the scoring of true-false tests. *J. Appl. Psychol.*, 1925, 9, 339-348.

*University of Minnesota
Minneapolis, Minnesota*

THE RELATIVE INFLUENCE OF INTELLIGENCE AND SOCIO-CULTURAL STATUS UPON THE INFORMATION POSSESSED BY FIRST-GRADE CHILDREN¹

K. C. GARRISON

Problems of diagnosis and prognosis in the various phases of educational and psychological measurements are ever becoming more and more important, since advancements in the fields of learning and guidance go hand in hand with the development of better means of diagnosis and prognosis. Recent years have witnessed a growing knowledge and interest in the different types of tests. A review of the articles devoted to this work, as presented yearly in the *Psychological Bulletin*, will reveal further this growing interest.

In approaching the problem the relative influence of intelligence and socio-cultural status upon the information of first-grade children, the writer does not propose this as a problem wholly synonymous with that of nature *vs.* nurture, as studied by Terman (7), Kelley (5), Freeman (2)

¹Read at the twenty-sixth annual meeting of The Southern Society for Philosophy and Psychology, April 3d and 4th, 1931, University of Virginia.

and others. Still, it is generally held that intelligence is somewhat comparable to nature, while socio-cultural is comparable to nurture. That intelligence as measured by intelligence tests is wholly a result of heredity is questioned by certain investigators who have found the intelligence quotient to vary from 20 to 30 points over gains made by control groups. Burks concludes from a careful study published in Part I of the *Twenty-Seventh Yearbook of the National Society for the Study of Education*:

"The total contribution of heredity (i.e., of innate and heritable factors) is probably not far from 75 to 80 per cent.

"Measurable environment one standard deviation above or below the mean of the population does not shift the I.Q. by more than 6 to 9 points above or below the value it would have had under normal environmental conditions.

"The maximal contribution of the best home environment to intelligence is apparently about 20 I.Q. points, or less, and almost surely lies between 10 and 30 points. Conversely, the least cultured, least stimulating kind of American home environment may depress the I.Q. as much as 20 I.Q. points. But situations as extreme as either of these probably occur only once or twice in a thousand times in American communities" (1, pp. 308-309).

Again, one cannot say that socio-cultural status is wholly a matter of environment and thus unrelated to factors of heredity. The various studies pertaining to the economic, social, cultural, and educational background of children with different IQ levels will reveal this definite and positive interrelationship to exist. Even though it is practically impossible to separate wholly intelligence and socio-cultural status of our subjects, this fact will not to any large degree invalidate the value and reliability of the data thus gathered.

The subjects for this investigation were selected from four sections of the first grade in the Raleigh Public Schools. One hundred and fifty-six subjects were given the Goodenough Drawing Scale for Measurement of Intelligence (4). One would expect a high correlation between the ordinary fact-finding intelligence test and information, since they are measuring to a large degree, the same thing. The Goodenough test was chosen because it is a performance test, and furthermore because it is different in nature from the information test used. Goodenough reports a reliability coefficient of $.937 \pm .006$ between the original scores on this test and scores earned by a retest on the following day (4, p. 48). Moreover, she reports a correlation of .74 between the Stanford-Binet IQ's and her Drawing IQ's. These subjects had all been in school five weeks prior to taking the test. None of these subjects used in this investigation had had any specialized type of training in drawing. Therefore, in so far as ideal subjects for this test is concerned, these data should be quite reliable and valid.

A socio-cultural scale was devised by combining a questionnaire used in gathering data from the homes represented in a study (by R. M. Elliot and D. G. Paterson of the University of Minnesota) on the relationship of certain environmental factors to measures of mechanical ability (3), and a questionnaire being used in the Raleigh Schools for gathering data pertaining to the home life and conditions of school children. The scale thus devised consists of 34 items, each item having the value of one point.

The Sangren Information Tests for Young Children (6) was used for determining the general quality and range of information possessed by these first-grade subjects. Sangren claims that in classifying children these Information Tests and the Stanford Revision of the Binet Test will agree about 95 per cent of the time, while it is claimed that the Stanford-Binet Test and the Information Tests will agree with the teachers' classification about 80 per cent of the time. Furthermore, by means of correlations involving the usual procedure he shows that this test meets this criterion for reliability and validity.

The drawing mental age and the socio-cultural status were found for these 156 first-grade subjects. These subjects were then paired first with respect to mental age, chronological age, and sex, but with a significant differential score for each pair on the socio-cultural scale. Thirty-eight pairs of subjects were thus found and each of these subjects was given the Sangren Information Tests. Again the subjects were paired with respect to socio-cultural status, chronological age, and sex, but with a significant differential score for each member of each pair in mental age. Thirty-three pairs of subjects were thus obtained. Twenty-nine of these 66 subjects were among the first list studied; therefore 37 more subjects were given the Information Tests, making 103 subjects to whom the Information Tests were given.

Table 1 presents the average information scores for the two groups studied. An analysis of these results will show that Groups A and B which were paired with respect to chronological age and socio-cultural were rather accurately paired, the difference between the average socio-cultural status for the two groups being only .21 of a point or a matter of only .037 of a probable error on the total socio-cultural distribution. The average mental age difference for this group was 23.05 or only one month short of two years. When we recognize that this is a fairly homogeneous group of children in many respects (the range in chronological age was from 6 years to 7 years and 6 months for the 103 subjects), then this mental age difference becomes more significant. Analyzing the other group, which is divided in Table 1 into Groups C and D, we notice a difference in the socio-cultural status of 16.93 points. The subjects of this group were paired with respect to mental age, sex, and chronological age. The average difference of .1 of a month in the average mental age in favor of Group C is too small for any statistical consideration. This table (Table 1) reveals

TABLE 1

THE AVERAGE INFORMATION SCORES FOR THE VARIOUS GROUPS OF FIRST-GRADE CHILDREN

Group	Information score
A. Subjects with an average socio-cultural status score of 15.36, but with an average mental age of 82.8 months	134.7
B. Subjects with an average socio-cultural status score of 15.15, but with an average mental age of 59.75 months	124.1
C. Subjects with an average socio-cultural status score of 23.63, but with an average mental age of 73.4	141.1
D. Subjects with an average socio-cultural status score of 6.7, but with an average mental age of 73.3 months	118.4

very clearly the accuracy of the pairing of the two groups with respect to an extremity and similarity of the two variables studied as they relate to information.

The average difference on the information test between Groups C and D was 22.7 points against a difference of 10.6 points between Groups A and B. Since Groups A and B were differentiated on the basis of mental age and Groups C and D on the basis of socio-cultural status, one would conclude from this that a superior socio-cultural is more important in developing information in the preschool child than is mental development. From the data one would infer that the mental age difference of 23.05 months between groups A and B is the same as a difference of 16.93 points between the socio-cultural status of Groups C and D. In order to ascertain the reliability of such an inference, these differences were studied in terms of probable error units for the entire group. Table 2 presents both the total differences in the information test scores between the two pairs of

TABLE 2

AVERAGE DIFFERENCES BETWEEN THE TWO PAIRS OF GROUPS ON THE SANGREN INFORMATION TESTS

Total average difference between Groups A and B	10.6
Total average difference between Groups A and D	22.7
Total average difference in mental age probable error units (Groups A and B)	1.44
Total average difference in socio-cultural probable error units (Groups C and D)	3.94

groups as well as the difference in terms of probable error units for the variables studied.

This latter treatment in terms of probable error units of the variables concerned gives a difference of 2.74 times as much for the groups differentiated on the basis of socio-cultural status as for the groups differentiated on the basis of mental age. By placing the difference in terms of probable error units for the two factors studied with reference to information, the units are more nearly equalized with respect to area of difference in terms of the total distribution.

In order to further check the results obtained by the equivalent group method the entire group of 103 subjects were thrown into one group and the three variables correlated. Table 3 presents the results of these correlations. These results indicate that socio-cultural status as measured by the scale here devised is more closely related to the information possessed by first-grade children than is drawing mental age.

TABLE 3

CORRELATIONS OF INTELLIGENCE AND SOCIO-CULTURAL STATUS WITH THE
INFORMATION POSSESSED BY FIRST-GRADE CHILDREN

($N = 103$)

Socio-cultural status and information.....	.425
Drawing mental age and information.....	.237

It has already been stated that the subjects were paired, first, with respect to equivalent mental ages, sex, and chronological age but differentiated on the socio-cultural scale, and, secondly, with respect to socio-cultural advantages, sex, and chronological age but differentiated on the mental age drawing scale. Thus, we have here a process of selection and elimination in which those subjects dissimilar in both intelligence and socio-cultural status tended to be eliminated. This, of course, would tend to render any correlation between intelligence and socio-cultural status somewhat valueless in so far as a true measurement of relationship was concerned. By the multiple correlation technique the correlation between the two variables combined and information was raised to .54.

The writer does not contend that this investigation has measured intelligence or socio-cultural status fully. Certainly, one can conclude that these tests measure these elements in part, and, in so far as they measure these factors, one would conclude that information of first-grade children is related to each of these but more closely related to socio-cultural status. The relationship between information and socio-cultural status is here found to be from two to three times as great as that between drawing mental age and information. With the inclusion of further factors into these measurements, one would expect a higher correlation. One cannot generalize from this study made with first-grade children that more mature subjects would give the same results. Certainly such factors as specialized training,

maturity, social institutions, customs, and traditions would each have its influence in affecting the relationships thus found.

REFERENCES

1. BURKS, B. S. The relative influence of nature and nurture upon mental development: a comparative study of foster parent—foster child resemblance and true parent—true child resemblance. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 219-317.
2. FREEMAN, F. N., *et al.* The influence of environment on the intelligence, school achievement, and conduct of foster children. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 102-217.
3. ANDERSON, L. D. The relationship of certain environmental factors to measures of mechanical ability. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 2, 1928, 137-150.
4. GOODENOUGH, F. L. The measurement of intelligence by drawings. Yonkers, N. Y.: World Book Co., 1926. Pp. xi+177.
5. KELLEY, T. L. The influence of nature and nurture upon native differences. New York: Macmillan, 1926. Pp. vii+49.
6. SANGREN, P. V. Sangren Information Tests for Young Children. Yonkers, N. Y.: World Book Co., 1930. Pp. iii+28.
7. TERMAN, L. M. Genetic studies of genius: Vol. 1. Mental and physical traits of a thousand gifted children. Stanford University, Calif.: Stanford Univ. Press, 1925. Pp. xv+625.

North Carolina State College
Raleigh, North Carolina

SUGGESTED CRITERIA FOR WRITING ATTITUDE STATEMENTS

CHARLES K. A. WANG*

The measurement of attitude by using attitude scales has in the past two or three years attracted wide-spread interest. Investigators in many fields are not only endeavoring to measure attitudes with ready-made scales but are also constructing various scales of their own. The successful construction of an attitude scale, however, requires more than knowing merely the mechanical steps of the procedure. In fact, the crucial task which first confronts one in constructing an attitude scale for a given issue is the collection of attitude statements on that issue. This is not a technical part of the construction method, but the success or failure of the scale depends much upon how well the initial list of statements is compiled and edited.

The purpose of this paper is to suggest a number of practical rules for writing attitude statements. These suggestions naturally are of special interest to those who construct attitude scales.

*Thanks are due the authors of several attitude scales for the use of certain illustrative statements. I am especially indebted to Prof. L. L. Thurstone through whom I obtained the original sorting data on these statements.

Aside from the fact that care must be taken to cover adequately the entire range of attitudes so as to prevent a break in any part of the scale and to include a variety of arguments so as to add reliability to the scale, the statements must be edited with regard to sentence construction and choice of words. The criterion of ambiguity and that of irrelevance (1, p. 44), valuable as they are in the selection of statements for the final scale, should not be depended upon too much in compiling the initial list of statements. Much time can be saved and statistical labor avoided if the experimental list does not contain too much worthless material.

The criteria suggested below are by no means to be taken as final but have been compiled from work done on statements for a series of attitude scales. They may be considered as an elaboration of the informal criteria given in Thurstone and Chave's monograph. Whenever possible, actual sorting data are presented in order to illustrate the ambiguous effect of violating the rule under discussion. Since normally a good attitude statement is distributed in no more than three or four consecutive piles, the significance of these sorting data is apparent.

1. An attitude statement must be debatable. It must represent only an opinion which has no general acceptance. Thus, a universal truth or a statement of fact should never be used in an attitude scale. To illustrate:

Bad: Unions are organized to protect labor.

Better: Unions are desirable for protecting labor.

The first statement is one with which everyone would agree. Its endorsement indicates not any kind of an attitude but the acknowledgment of a fact. If it were included in an attitude scale, it would be endorsed by people who are opposed to unions as well as by those who are friendly to them. For a similar reason:

Bad: It is hard on the children to have the mother working.

Better: Women with children should not work.

2. All statements on a given issue should belong, as nearly as can be judged, to the same attitude variable. That is, they must be not only relevant to the issue but belong to the linear continuum that is being measured. As violations of this rule constitute the most common type of faulty statements, it is well that every attitude statement be considered in terms of this criterion. Examples:

Statement: In an ideal society there would be no law. (From a scale on attitude toward law, where the variable being measured is from complete respect to utter disrespect for law.)

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII
Frequency:	5	6	11	10	7	10	3	2

Statement: We are governed by laws, not by men. From the same scale as above.)

Sorting data:

Pile:	II	III	IV	V	VI	VII	VIII	IX	X	XI
Frequency:	1	3	5	1	3	0	15	15	10	1

Statement: Total abstinence from liquor can be accomplished only by education. (From a scale on attitude toward prohibition, where the variable being measured is from complete approval of, to violent opposition to, prohibition.)

Sorting data:

Pile:	II	III	IV	V	VI	VII	VIII	IX	X
Frequency:	1	8	3	6	9	8	4	1	1

The wide scatter in each case shows that the statement was ambiguous with reference to the principal issue.

3. An attitude statement must not be susceptible to more than one interpretation. It must contain no word or phrase which can be construed to mean different things by different individuals. A few illustrations follow:

Statement: Birth control legislation is a disgrace to our civilization.

Sorting data:

Pile:	I	II	III	—	—	—	—	IX	X	XI
Frequency:	10	19	9					4	13	7

The ambiguity of this statement is apparently caused by the fact that the statement can be endorsed by people who oppose as well as by those who endorse birth control, although with different interpretations.

Statement: Anyone who makes others suffer should suffer in return. (From a scale on attitude toward the treatment of criminals.)

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Frequency:	5	7	1	1	3	1	2	5	13	13	3

The ambiguity here seems to lie in the possibility that the term "anyone" may refer to criminals or to society.

Statement: There can be no compromise with the evolutionists.

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII	IX	X
Frequency:	24	7	7	4	3	1	1	0	2	1

It is interesting to observe the consistency of the judgments by noting the results on a counter statement as shown below:

Statement: There can be no compromise with the enemies of evolution.

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Frequency:	2	3	1	0	2	0	2	5	5	11	18

To endorse one of these statements does not clearly show to which end of the scale the subject belongs because the statements can be interpreted in different ways.

4. Avoid "double-barreled" statements. While it is sometimes necessary, in order to express neutrality or a mild attitude, to balance one idea with an opposite idea, experience has shown that this type of statement usually causes either high ambiguity or a break at the neutral pile, i.e., resulting in a bimodal distribution. To illustrate:

Statement: Athletic conditions are bad, but officials are trying to improve them.

Sorting data:

Pile:	III	IV	V	VI	VII	VIII	IX
Frequency:	2	7	19	19	42	9	2

Statement: Birth control would be all right if we could prevent people from taking immoral advantage of it.

Sorting data:

Pile:	IV	V	VI	VII	VIII
Frequency:	2	17	2	39	3

These distributions show that the subjects had difficulty in deciding whether the real attitude is favorable or unfavorable to the issues involved.

5. An attitude statement should be short. It should rarely exceed fifteen words in length. Most statements should be much shorter. A long statement can usually be reduced to a shorter one without altering its essential point. For example: Instead of saying,

"The Bible represents the sacred word of God and should be respected by everyone as such"

the statement can be reduced to merely,

"The Bible is the sacred word of God."

In writing attitude statements, it is well to try to shorten the length of each sentence written. In doing so, one usually also avoids the violation of many of the other rules here mentioned.

6. Each attitude statement should be complete in denoting a definite attitude toward a specific issue. Do not assume that the issue in question can be understood without specific reference to it. Thus, do not use "laws" to mean prohibition laws, or "it," "they," etc., unless the reference is perfectly clear *within* the statement in which they are used.

7. Each attitude statement should contain only one complete thought. Too many ideas in one statement cause confusion in interpreting the attitude and thus increase the chance of high ambiguity. The following example illustrates the effect of violating this rule:

Statement: Prohibition laws and enforcement are merely political games and there is no moral issue in keeping or breaking them.

Sorting data:

Pile:	IV	V	VI	VII	VIII	IX	X	XI
Frequency:	3	27	129	66	30	27	12	6

Another example of violation of this rule is the following:

"The church was established to serve a useful purpose but it has outlived its time; therefore, it is doing more harm than good."

The remedy for this type of statement is to break it up into two or more shorter statements, thus:

- a. The church serves a useful purpose.
- b. The church has outlived its usefulness.
- c. The church does more harm than good.

8. Avoid grouping two or more complete sentences as one attitude statement. Do not transplant quotations by the paragraph *en bloc*, but rewrite

them into one single sentence or several separate statements. For example, instead of quoting:

"War is the concentration of all human crimes. Under its standard gather violence, malignity, rage, fraud, perfidy, rapacity, and lust. If it only slew men it would do little. It turns man into a beast of prey."

as one attitude statement, it would be better to rewrite this into three separate statements, thus:

a. War is the concentration of all human crimes.

b. War is degenerating.

c. War makes man a beast of prey.

9. An attitude statement should be clear-cut and direct. Avoid statements which are not directly an attitude but from which an attitude is to be inferred, unless the inference is clear and unquestionable. To illustrate, as a statement of attitude on patriotism, the following is ambiguous as the sorting data demonstrate:

Statement: The makers of American policy are too idealistic and should not get mixed up in European affairs.

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII	IX
Frequency:	5	2	6	4	5	5	7	5	6

The wide scatter indicates that the subjects tried to guess at the attitude of the statement. Illustrating further, the following statements, as attitudes toward war, are likewise too indirect to be satisfactory:

"Military instruction should be separated from colleges and universities."

"Patriotism is the imperative end of education."

"Loyalty to our country comes before world-brotherhood."

On the other hand, some of the best attitude statements are those which infer unmistakably an attitude but do not explicitly assert it. Examples of this type are "Might is right" as an attitude favorable to militarism, and "We do not need more babies but better ones" as an attitude favorable to birth control.

10. Use with care and moderation such words as "only," "mere," "just" (in the sense of only), "merely," etc. Many statements containing one or another of these words have been found to cause ambiguity or bimodal distribution, even though on deliberation the affect may be clear. Examples:

Statement: Only by taking the money out of football can it be made really amateur.

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII	IX	X
Frequency:	6	16	11	20	12	2	11	2	2	1

Statement: In my thinking God merely means an ideal.

Sorting data:

Pile:	II	III	IV	V	VI	VII	VIII
Frequency:	1	4	17	19	4	25	7

Statement: I respect only those laws which represent the wishes of the majority.

Sorting data:

Pile:	II	III	IV	V	VI	VII	VIII
Frequency:	3	4	12	11	0	22	2

These distributions show that by qualifying a statement with such words as "only," "mere," etc., it becomes difficult to interpret the attitude.

11. Avoid colorless expressions or statements lacking affect. While it is not necessary that every attitude statement must be emotionally toned, it should always represent some clearly defined conviction. Thus, do not say:

"The unions (or anything else) are all right."

"Churches are beautiful buildings." Etc.

To illustrate further:

Statement: Intercollegiate athletics are a necessary evil.

Sorting data:

Pile:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Frequency:	2	4	5	8	31	4	34	5	5	1	1

Since this type of statement expresses no conviction in one way or another, it can be endorsed by a person who may belong to any part of the scale.

12. Whenever possible, write an attitude statement in the form of a simple rather than a complex or compound sentence. The simple kind of statement reduces the chance for a wrong interpretation. Thus:

Bad: Women have always had enough to do to look after their men, and they should not need more.

Better: Women work enough without outside employment.

13. When a statement cannot be made in the form of a simple sentence, write it as a complex rather than a compound one. For example:

Awkward: Some code of ethics is necessary for the guidance of our conduct and the Bible may well serve that purpose.

Better: Since we need an ethical code for guiding conduct, the Bible is indispensable.

Best: The Bible is necessary for guiding conduct.

14. It is usually better to use the active rather than the passive voice. Thus,

"Punishment does not deter crime"

is preferred to

"Crime cannot be deterred by punishment."

An exception to this rule, however, will be discussed in the next paragraph.

15. In general, use the term of the issue as the subject of a statement. This is desirable in order to secure proper emphasis and attention. Hence it is permitted even in violation of Rule 14. For example, if the issue is on attitudes toward public office, it is better to say, "Public officials are controlled by crooks," than to say, "The crooks control public officials." On the other hand, if the issue is on attitudes toward the social influence of crime, it would be better to use the second statement in preference to the first.

16. Avoid high-sounding words, uncommon words or expressions, technical terms not ordinarily understood, etc. When a scale is being prepared for use in a specific age, school, or sociological group, the vocabulary of

that group should be borne in mind. In any case, it is better to write statements in the simplest and most precise language possible.

In addition to the foregoing criteria, there may be mentioned several general rules, based largely upon good usage in English. These rules improve sentence structure although they are not necessarily concerned with the scale values or the Q-values of the statements.

1. Avoid a negative expression whenever a positive one can be substituted. Thus, use "disagree" instead of "not agree" "difficult" instead of "not easy," etc. Exceptions, of course, are permitted when the negative effect is desired.

2. Avoid double infinitives, especially in a short statement. For example, instead of saying, "To work on Sunday is to be immoral," say, "Working on Sunday is immoral." Usually, in a case of double infinitives, at least one of them can be changed into a present participle.

3. Do not use redundant phrases. To illustrate:

Bad: We should not knock but boost our public officials.

Better: We should boost our public officials.

Bad: Communism should be absolutely banished at all costs.

Better: Communism should be banished absolutely.

By eliminating the redundancy, these statements are improved by their shorter length, yet lose none of their effectiveness as attitudes.

4. Avoid excessive use of such phrases as "I think that . . ."; "I believe that . . ."; "I feel . . ."; etc., to precede a statement. Simply make the direct statement. Such prefaces are excess baggage.

5. Avoid double negatives. Statements such as the following have sometimes been found to result in high ambiguity:

"I don't believe in disobeying the law."

"I do not dislike the negroes."

This double negative type of statement lacks the force of directness with which every attitude should be expressed.

In conclusion, the distinguishing feature of an attitude statement lies mainly in that it expresses an attitude. To be sure, the technical requirement of the statement is that it belongs to the attitude continuum being measured, but that is adequately provided for by objective criteria. Otherwise, an attitude statement differs little from other statements, at least so far as language structure is concerned. It should therefore follow all the rules of good English (except occasionally for special sociological groups), with the added feature of being simple but precise, short but complete. The criteria suggested in this paper are discussed in detail only because it is hoped that, in doing so, they would help in obtaining initial statements that are more likely to produce satisfactory attitude scales.

REFERENCE

1. THURSTONE, L. L., & CHAVE, E. J. The measurement of attitude. Chicago: Univ. Chicago Press, 1929. Pp. 96.
The University of Chicago
Chicago, Illinois

RACE, SEX, AND CLASS DIFFERENCES IN ABILITY TO ENDURE PAIN

CHARLES C. JOSEY AND CARROLL H. MILLER

In an effort to discover whether there are race, sex, or class differences in ability to endure pain with courage, the writers sent the following questionnaire to 120 physicians and dentists:

1. Check those of the following races or nationalities of which you have more than an occasional case:

- | | | |
|------------|------------|------------------------|
| 1. Negroes | 5. Irish | 9. Scandanavians |
| 2. Greeks | 6. Germans | 10. Indians (American) |
| 3. English | 7. Scotch | 11. Italians |
| 4. Chinese | 8. Dutch | 12. Jews |

2. Have you observed that some races show greater self-control than others while enduring pain? (Yes or No)

3. If so, arrange a list of races in order of their ability for self-control in pain, beginning with those that show the greatest self-control. List only those races with which you have actually had experience.

4. Have you noticed any sex difference in ability to stand pain? (Yes or No)

5. If so, which sex shows the greater ability?

6. Have you noticed any class difference in ability to endure pain? (Yes or No)

7. If so, which class (the well-to-do or the laboring) shows more self-control?

We received 60 replies. In addition to answering our questions, a number gave emphasis to their answers in a letter or in a note written on the questionnaire. Two of these are selected because of their seeming objective character. One stated that following a major operation on well-to-do whites catheterization of the urinary bladder is usually necessary, though this is seldom necessary with Negroes. Another stated that he has noticed that following injuries shock is least severe in Scandanavians and much exaggerated in Italians and Jews.

Race Differences. Of the 60 who replied, 12 stated that there are no race differences; 48 stated that there are, but, as one failed to rate the different races, no use could be made of his judgment in our calculations. As the doctors were asked to rate only those races with which they had had experience, the number rated by the different doctors naturally varied. The modal number rated was six. We decided to make that our basis for estimating how much weight should be attached to the opinions of men of varying breadth of experience. If a person rated six the weight accorded to his judgment was 21, the sum of the digits from six to one; if he rated nine the weight became $9/6$ of 21 or 31.5; if he rated three the weight became $3/6$ of 21 or 10.5, and so on for all the others. The weight given

the race rated first by a doctor who rated nine was 9/45 of 31.5 or 6.3; the race rated second was given a credit of 8/45 of 31.5; and so on; the race rated first by a doctor who rated three was 3/6 of 10.5 or 5.25; the one rated second was 2/6 of 10.5, and so on.

Table 1 sets forth the most interesting results.

TABLE 1

Race	No. of opinions	Mean score	Rank	Mean deviation
Indian	20	4.41	1	1.49
German	35	4.38	2	1.19
Dutch	16	3.83	3	.99
Negro	12	3.64	4	1.67
Scotch	16	3.61	5	1.10
Irish	31	3.47	6	.98
English	29	3.45	7	1.35
Scandinavian	32	3.44	8	1.65
Greek	9	2.72	9	1.16
Italian	9	2.52	10	1.00
Jew	19	1.37	11	.53

Note: No statement is made regarding the Chinese as only two of our informants reported on them.

If we divide the Europeans into two groups, the North Europeans and the South Europeans, we get the rather significant results shown in Table 2.

TABLE 2

Race	Median Score	Rank
North European	3.70	2
South European	2.63	4
Jew	1.37	5
Indian	4.41	1
Negro	3.64	3

Admitting the roughness of the above classification as far as the Europeans are concerned, yet the differences here indicated between races seem to the writers sufficient to suggest that here is a field of race psychology that should be studied in a more reliable manner.

Sex Differences. Of the 60 who replied four did not answer the question regarding sex differences, and eight stated that there are no sex differences. Since those who did not reply probably have no decided views, we cannot go far wrong if we add them to those who said there are none. On this basis our results are as follows:

Number who believe there are no sex differences	12, or 20%
Number who believe the female is superior	42, or 70%
Number who believe the male is superior	6, or 10%

Class Differences. Forty-one, or approximately 70%, of those who returned the questionnaire stated that the laboring class is superior; eight, or approximately 13%, stated that the well-to-do class is superior.

*University of South Dakota
Vermilion, South Dakota*

TESTING SOME ASPECTS OF PERSONALITY

EDWIN G. FLEMMING

Personality is a very complex affair. It is difficult to define and seems to comprehend many factors. So-called tests of personality really do not test personality as such, but rather pretend to measure only certain aspects of personality.

While personality is generally regarded as a whole by the layman and the Gestalt psychologist, nevertheless it seems desirable to try to analyze it into its parts, or into its supposed parts, in order to attempt to understand what personality really is.

Obviously the pleasingness of an individual to his fellows is an aspect of personality; it is likewise generally considered quite as obvious that the emotional expressiveness of the individual, his emotional balance, and his general social adjustment are parts of or expressions of his personality. In this study we have taken these four things as our criteria.

In the field of tests, the Colgate Schedule C 2, a test of introversion and extroversion, is considered a test of personality. Woodworth (4, pp. 555-556) maintains that intelligence is one of the factors in personality, and with that assertion we are not likely to quarrel since it seems reasonable that it should be. Since social adjustment is one of the aspects of personality, it would be supposed that the social intelligence of a person as measured by such a test as the Social Intelligence Test of the George Washington University Series would show itself in some of the aspects of personality. And, further, it would seem that a test of the home background would throw some light upon personality problems, since those individuals coming from superior homes would reasonably be expected to be able to fit in better in a complex environment, and to make a better adjustment to a multiplicity of social demands.

In the field of tests, then, we have chosen the Colgate Schedule C 2, the George Washington University Social Intelligence Test, the Sims Score Card for Socio-Economic Status, and the Army Alpha Intelligence Test. We have also chosen the Thorndike Examination for High School Graduates, because, while it is not claimed to be an intelligence test, it nevertheless measures one aspect of intelligence as reflected in college grades.

The subjects used were the students taking the elementary course in psy-

chology at a small college. There were about two hundred men and women; but, due to some difficulties in administration, absences of different students on different days, not always is the same number of students used in each correlation. The students were all of sophomore rank or above, but preponderately sophomores. The students knew each other very well, and had many opportunities in the small college town of about twenty-five hundred inhabitants to see each other in all sorts of situations.

The scores on the criteria were secured on a rating sheet. Mimeographed sheets containing the names of all the students, and opposite each name the figures from one to ten, were distributed. One set of names for each of the criteria was given out. The instructions defined the criteria.

On the rating sheets for pleasingness of personality, for example, the directions were as follows:

Rate the following individuals on the basis of how they affect you. Do not consider their general reputations. Try to answer the question: "Is my response to this individual pleasant or unpleasant?" If your response is very pleasant, draw a circle around the figure ten (10); if your response is unpleasant in the extreme, put a circle around the one (1). In order to gauge the pleasantness of your response, consider the response that you make to the individual you like best as the standard for circling the ten (10). In gauging the unpleasantness of your response, consider the response you make to the individual you dislike the most as the standard for encircling the one (1). Rate the rest with relation to these two as extremes. **DO NOT RATE ANY INDIVIDUAL WHOM YOU FEEL YOU DO NOT KNOW WELL ENOUGH TO RATE.**

Similar instructions were given for the ratings on emotional steadiness, emotional expressiveness, and adjustment to social situations.¹ The number of ratings secured on the subjects ranged from 3 to 134; in no case, however, was a subject used unless he had at least five ratings on the particular trait under consideration. The average number of ratings on pleasingness was about 51, on steadiness about 35, on expressiveness about 34, and on social adjustment about 37.

In pleasingness of personality, in expressiveness, and in adjustment a high score means that the individual is more pleasing, more expressive, or better adjusted than his fellows. But in steadiness a high score means that the subject is less steady, or more unsteady than his fellows. This needs to be borne in mind in interpreting the correlations.

A description of the tests used is hardly necessary since they are all well known and frequently used, with the possible exception of the Sims score

¹For a more detailed discussion of the criteria, definitions of the criteria, and the number of ratings see (2).

TABLE 1

SHOWING CORRELATIONS BETWEEN PLEASINGNESS OF PERSONALITY AND THE TESTS INDICATED

Men only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .10	74
Army Alpha	.12	81
Thorndike	.02	62
Sims score card	— .09	81
Social Intelligence	.32	77

card. It asks the usual questions about the number of rooms in the home, the number of books, the number in the family, the number of servants, the manner of spending leisure, the custom with respect to vacations, and so on. It has been quite well standardized, however, and has a reliability coefficient of .94 (1).

In treating the data it has seemed wise to handle the men and women separately, so as to discern, by the way, whether there were any sex differences. However, the scores on the criteria represent the judgments of both men and women.

Table 1 shows the correlations between the criterion, pleasingness, and the tests indicated, with the men as subjects.

While these results do not throw any doubt upon the validity of the Colgate Schedule, the test of introversion, they do, on the other hand, corroborate the statement of Laird (3, p. 315) to the effect that there is no choice between introversion and extroversion on the basis of desirability. It appears quite clearly that introversion or extroversion has no significant relation to pleasingness of personality.

Likewise, while this study does not show that intelligence is not a factor in personality, it does show that intelligence, measured either by Army Alpha or by the Thorndike examination, is not a significant factor in determining the pleasingness of one's personality. However, the question may be raised: Just what is personality, and what is the relation of intelligence to that concept of personality?

It further appears that the nature of the home environment has no bearing upon the pleasingness of the individual.

Social intelligence, on the other hand, seems to have some bearing upon the pleasantness of the personalities of men. This is a fairly significant finding, although the correlation is low. It nevertheless is high enough to show a definite tendency, and to warrant further investigation along that line.

The next table, Table 2, shows the correlations between these tests and the steadiness of the emotional response of men. The question put in the

TABLE 2
SHOWING CORRELATIONS BETWEEN STEADINESS OF EMOTIONAL RESPONSES AND
THE TESTS INDICATED
Men only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .15	74
Army Alpha	— .01	81
Thorndike	— .03	62
Sims score card	.22	81
Social Intelligence	— .06	75

directions for this rating sheet was: "Is he erratic; is it impossible to tell how he will respond emotionally to any situation?" The high score represents the most erratic or least study.

It now appears that emotional steadiness, or unsteadiness, is not indicated by any of the tests used. If emotional steadiness is legitimately a part of personality, or one aspect of it, then none of the tests, or the things they measure, have any relation to such an aspect of personality.

Table 3 shows the correlations between these same tests and emotional expressiveness of men. Here, again, we find no significant relationship showing itself. And again we may observe that if emotional expressiveness is an aspect of personality, then intelligence, introversion or extroversion, home environment, nor social intelligence have any relation to personality from this angle.

When the subjects were asked to make ratings on adjustment, they were instructed that "well adjusted" means that he "fits in well" into the various situations of life; in social situations as well as in the academic environment; with old people as well as with people of his own age; with individuals of the opposite sex as well as his own sex; with people of importance as well as with those of not much importance; and so on. "Poorly adjusted" means that he does not "fit in well"; that he is "queer"; that he appears to be uncomfortable in all or some of these situations.

TABLE 3
SHOWING CORRELATIONS BETWEEN EMOTIONAL EXPRESSIVENESS AND THE TESTS
INDICATED
Men only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .11	74
Army Alpha	.15	81
Thorndike	.11	62
Sims score card	.25	81
Social Intelligence	.25	75

TABLE 4
SHOWING THE CORRELATIONS BETWEEN SOCIAL ADJUSTMENT AND THE TESTS
INDICATED
Men only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .02	74
Army Alpha	.21	81
Thorndike	.13	62
Sims score card	.21	81
Social Intelligence	.46	75

In Table 4 the correlations between social adjustment and the tests used are shown. Here again none of the tests shows any significant relation to the criterion except social intelligence, which shows a positive tendency to be associated with adjustment in the case of men.

While it appears that among the tests used only social intelligence shows any relationship of any consequence with any of the four criteria—pleasingness, steadiness, expressiveness, and adjustment—one would not be warranted in concluding that neither intelligence, introversion-extroversion, nor socio-economic status has anything whatever to do with personality. Personality is a complex affair, and the criteria represent only four possible aspects of personality. It may very well be that intelligence and introversion-extroversion are themselves aspects of personality of equal value with the four criteria used. The most we can say is that of all the measures used only social intelligence appears to be related to pleasingness of personality and to social adjustment of college men.

It is important to emphasize the fact that college men were the subjects, because it is quite possible that with a larger and less selected group, with a more random sample of the entire population, a closer relationship might be found.

And now let us turn to the study of the women. Table 5 shows the correlations between the criterion, pleasingness, and the tests used.

With the women we find that none of the tests shows any significant rela-

TABLE 5
SHOWING THE CORRELATIONS BETWEEN PLEASINGNESS AND THE TESTS INDICATED
Women only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .04	112
Army Alpha	.13	118
Thorndike	— .15	83
Sims score card	— .13	111
Social Intelligence	.08	42

tionship to pleasingness of personality; not even social intelligence, which has a correlation of .32 for the men. That the absence of any more definite relationship with the women is due to the relatively few cases (only 42 as compared with 77 for the men) is entirely possible, but we are inclined to doubt that it is probable. If it is true that social intelligence, or a knowledge of what is right and proper in social situations as indicated by the George Washington Social Intelligence Test, is of more importance to men than to women in contributing to a pleasing personality, that fact is contrary to the generally accepted notion that it is more important for women to be familiar with social conventions. If it is true, then it may mean a change in emphasis in the social education of men for the purpose of developing more pleasing aspects of personality.

It might occur to one that the reason for the difference is that the women in general show a higher social intelligence than the men, in which case the greater significance for the men would be understandable. The averages, however, show for the group of men studied a mean score of 99.8, and for the women a mean score of 95.6. So, for these two groups, the women certainly are not generally more intelligent socially than the men. Consequently, our first conclusion must stand, that for college men in this study social intelligence is of more importance as a contributing factor to a pleasing personality than it is for the women.

There is also another difference between men and women which is not so clearly revealed by the simple correlations; but you will notice that with the men the correlations between pleasingness and Army Alpha and between pleasingness and Thorndike are both low but both positive. With the women, however, the correlation between pleasingness and Army Alpha is positive, while the correlations between pleasingness and Thorndike is negative. Both correlations are relatively low; but the fact that the correlation between Army Alpha and Thorndike is high and positive, being .73, would suggest that something might be revealed by the method of partial correlation. When we hold Army Alpha constant the correlation between pleasingness and Thorndike becomes $-.36$, showing a relationship of definite tendency and of some probable significance.

Of course, the Thorndike examination is not claimed to be a test of general, native intelligence; whereas the Army Alpha makes some such claim. The Thorndike test probably measures intelligence plus acquired knowledge gathered either through superior industry or because of wider opportunity in the form of greater educational advantages. The virtual effect then of partialing out Army Alpha is to leave as a remainder a measure of greater industry or superior opportunity. In the college at which the subjects were secured, the girls came from a relatively restricted area, mostly from within the borders of the state. Practically all of them had attended the public schools. The home environment might have made a difference in the educational opportunity, but the correlation between

TABLE 6
SHOWING CORRELATIONS BETWEEN EMOTIONAL STEADINESS AND THE TESTS
INDICATED
Women only

Test	Correlation	Number of cases
Colgate Schedule C 2	.02	112
Army Alpha	— .13	118
Thorndike	.02	83
Sims score card	.19	111
Social Intelligence	— .10	42

pleasingness and the Sims score card shows a correlation of only —.13. So it is quite likely that educational opportunity has little to do with the significant trend shown when intelligence is partialled out. Besides, when Army Alpha is held constant the correlation between pleasingness and the Sims score card is raised to only —.16, still an insignificant relationship. Consequently, we may conclude that the partial correlation between pleasingness of personality and Thorndike with Army Alpha held constant shows the relationship between pleasing personality and academic industry. In that case those women who are academically industrious show a definite tendency to have not as pleasing personalities as the other women who occupy themselves with other matters.

Whether the industry is the cause of the lack in personality, or whether the lack of pleasingness is the cause of the industry as a kind of compensating device is an open question. It is quite possible that back of it all is an absence of beauty and taste in dress, or a lack of money with which to engage in social activities and entertainments. It would be of value in another study to get estimates of beauty along with measures of the other qualities which represent aspects of personality.

In Table 6 appear the correlations between emotional steadiness and the tests used in this study. They show no significant relationships.

In Table 7 the correlations between emotional expressiveness and the tests used indicate that the women coming from the better homes as measured by the Sims score card are more likely to be emotionally expressive than those coming from homes with lower ratings. At first glance it would seem that this is contrary to expectations, since ordinarily one would expect women from the better homes to be more moderate in the expression of their emotions; one would expect better control as a result of the cultural influences of the environment. However, it is quite likely that the emotional expression indulged in by college women is to a large extent in the form of enthusiasm and *joie de vie*. Where the emotion is in the form of anger it is probable that it is expressed because the individual from the better home knows more definitely what he wants, and is not adverse to

TABLE 7
SHOWING CORRELATIONS BETWEEN EMOTIONAL EXPRESSIVENESS AND THE TESTS
INDICATED
Women only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .09	112
Army Alpha	.08	118
Thorndike	.15	83
Sims score card	.34	111
Social Intelligence	.10	42

criticizing things when they do not go as they are supposed to go. Likewise, the probabilities are that those coming from the homes with lower ratings are not so sure of their social status, of their own feelings and ideas, and would be less likely to express what they feel among the democratic college group for fear that they may be wrong, and the disagreement might bring about feelings of inadequacy and inferiority. So it seems rather a normal social situation that the women coming from better homes should tend to be somewhat more expressive emotionally.

The correlation between expressiveness and the Sims score card for the men is .25, while the same correlation for the women is .34. With the men the relationship between expressiveness and socio-economic status seems to be of less significance than with the women. Perhaps the difference is due to chance; on the other hand, it may be due to the fact that women are generally more expressive than the men as was shown in another study (2). Since the women are more expressive on the average, the relationship existing with the quality of the home shows itself in a more pronounced manner.

Table 8 contains the correlations between social adjustment and the tests indicated. It shows no significant correlations. It is interesting that for women the relationship between adjustment and social intelligence is represented by a correlation of .19, while for the men the correlation is .53, the

TABLE 8
SHOWING CORRELATIONS BETWEEN SOCIAL ADJUSTMENT AND THE TESTS
INDICATED
Women only

Test	Correlation	Number of cases
Colgate Schedule C 2	— .01	112
Army Alpha	.16	118
Thorndike	.09	83
Sims score card	.14	111
Social Intelligence	.19	42

highest correlation secured in this study. For men, then, social intelligence is of far more importance as a contributing factor to their social adjustments than it is for women. As in the case of pleasing personality, the greater significance for men of social intelligence or knowledge is of some importance and may indicate a need for a changed emphasis upon social instruction for men.

CONCLUSIONS

Women coming from the better grade homes tend to be more expressive emotionally than those coming from homes of a lower socio-economic status. The correlation is .34. The same is not true of the men; the relationship between emotional expressiveness and socio-economic status being of no practical significance for them as indicated by the correlation of .25.

Women who are academically industrious tend to be less pleasing than others to their fellow students, since the partial correlation between pleasingness of personality and Thorndike with Army Alpha held constant is $-.36$. The factor of educational opportunity as a contributor is ruled out by the insignificant correlation, both simple and partial, between pleasingness and the socio-economic status, and the same with intelligence held constant. The simple correlation is but $-.13$, while the partial is only $-.16$.

Academic industry seems, on the other hand, to be of no significance to college men as a detractor from pleasingness of personality.

The most important finding, however, is that social intelligence or knowledge, as measured by the George Washington University Social Intelligence Test, is of more importance to men than it is to women, both with relation to a pleasing personality and to social adjustment. It is of more significance to men with respect to social adjustment than it is with respect to pleasingness to others. The relationship, and the significant difference of the importance of social intelligence for men and women, may indicate the need for a change in the social education of men.

REFERENCES

1. FLEMMING, C. W., & RUTLEDGE, S. A. The importance of the social and economic quality of the home for pupil guidance. *Teach. Coll. Rec.*, 1927, **29**, 202-215.
2. FLEMMING, E. G. Pleasing personality. *J. Soc. Psychol.*, 1932, **3**, 100-107.
3. LAIRD, D. A. The psychology of selecting men. (2nd ed.) New York: McGraw-Hill, 1927. Pp. 345.
4. WOODWORTH, R. S. Psychology. (Rev. ed.) New York: Holt, 1929. Pp. 590.

29 Claremont Avenue
New York City

BEST FRIENDS

EDWIN G. FLEMMING

Do birds of a feather flock together; or do opposites attract?

At the present time the answers to those questions remain matters of speculative opinion. To throw some scientific light upon them the present investigation was undertaken.

Various data were secured on about two hundred students taking a compulsory course in elementary psychology. They were mostly of sophomore rank in college. Among other things they were asked to indicate the name of their best friend of the same sex in college. The ratings and scores on various measures for each student were then compared with the ratings and scores on the same measures of the friend whose name was given, when that friend was among the subjects studied. Of course, if the friend was not included among the other subjects, the comparison could not be made.

The traits measured were as follows:

Pleasingness of personality.

Steadiness of emotional responses.

Emotional expressiveness.

Social adjustment.

Introversion-extroversion, by the Laird test.

Intelligence, by the Army Alpha test.

Intelligence, by the Thorndike examination.

Socio-economic status, by the Sims scale.

Social intelligence, by the George Washington University social Intelligence Test.

Scores for pleasingness, steadiness, expressiveness, and adjustment were secured by means of rating sheets. For pleasingness the directions were as follows:

Rate the following individuals on the basis of how they affect you. Do not consider their general reputations. Try to answer the question: "Is my response to this individual pleasant or unpleasant?" If your response is very pleasant, draw a circle around the figure ten (10); if your response is unpleasant in the extreme put a circle around the one (1). In order to gauge the pleasantness of your response, consider the response that you make to the individual you like best as the standard for circling the ten (10). In gauging the unpleasantness of your response, consider the response you make to the individual you dislike the most as the standard for encircling the one (1). Rate the rest with relation to these two as extremes. DO NOT RATE ANY INDIVIDUAL WHOM YOU FEEL YOU DO NOT KNOW WELL ENOUGH TO RATE.

TABLE 1

CORRELATIONS BETWEEN SCORES ON VARIOUS MEASURES BETWEEN INDIVIDUALS AND THEIR BEST FRIENDS

Trait	Men		Women	
	<i>r</i>	<i>N</i>	<i>r</i>	<i>N</i>
Pleasingness	.34	48	.53	61
Steadiness	.17	48	.20	61
Expressiveness	.15	47	.03	61
Adjustment	.38	48	.42	61
Laird	.56	39	.13	59
Army Alpha	.14	48	.11	60
Thorndike	.12	21	.51	45
Sims scale	.20	23	.24	45
Social Intelligence	.44	18	.54	11

Sheets containing the names of all the students taking the course accompanied the instructions, and opposite each name the figures from one to ten appeared. One set of names was given out with each sheet of instructions.

Similar instructions were given for the ratings on steadiness, expressiveness, and adjustment.¹ The number of ratings secured on the subjects ranged from 3 to 134; in no case, however, was a subject used who had less than five ratings on the particular trait considered. The average number of ratings on pleasingness was about 51, on steadiness about 35, on expressiveness about 34, and on social adjustment about 37.

In pleasingness, expressiveness, and adjustment, a high score means that the individual is more pleasing, more expressive, or better adjusted than his fellows. But in steadiness a high score means that the subject is less steady, or more unsteady than his fellows.

The tests used are all fairly familiar and require no detailed description.

In working up the results the men and women were treated separately, in order to see whether any significant sex differences might be revealed.

Table 1 shows the correlations found between the ratings and scores of the subject with the ratings and scores of his best friend.

These correlations indicate in general that there is a greater tendency for birds of a feather to flock together than for opposites to attract. In individual cases opposites may attract, but that none of the correlations is negative is probably of some significance. In no case, however, is the correlation so high that one can predict with any high degree of probability what the friend will be like.

¹For a more detailed discussion of the personality ratings, definitions of the personality traits, and the number of ratings, etc., see (1).

For the men the most significant correlation is the one on introversion-extroversion as measured by the Laird test. Introverts seek introverts as friends, while extroverts choose extroverts, as a fairly definite tendency. Pleasing individuals tend to have pleasing best friends. Those well adjusted tend to have friends who are well adjusted. And those of high social intelligence tend to associate with those who also have high social intelligence.

On the other hand, for the men there is no apparent relation between best friends on the basis of emotional steadiness, expressiveness, intelligence as measured either by the Army Alpha or the Thorndike examination, nor upon the basis of socio-economic status as measured by the Sims scale.

The average correlation for the women is somewhat higher than the average correlation for the men, indicating a tendency for women friends to be more alike than men friends. They are more alike than the men on pleasingness, adjustment, and social intelligence. But in introversion-extroversion they are less alike than men; women introverts apparently do not choose introverts as friends as the men very definitely tend to do. On that basis the selections of the women are more or less at random.

With respect to intelligence the situation is also different. If we consider the Army Alpha test a measure of native intelligence, and the Thorndike test a measure of native intelligence plus information acquired at high school, we see that neither men nor women choose their friends on the basis of similarity in native intelligence. But while men friends show no similarity in acquired information as measured by the Thorndike test, the women show a high degree of similarity; they correlate on Thorndike to the extent of .51.

It appears then, that in the case of both men and women, pleasing individuals, well adjusted to the social milieu, and of good social intelligence tend to select as their friends individuals similarly pleasing, well adjusted, and intelligent. It also appears that men are attracted by similarity in introversion, while women are not; but that women are attracted by similarity in acquired information, while men are not.

Another method of comparing the individual with his best friend is to compare the average scores for all the subjects with the average scores for all the best friends. Table 2 shows such average scores for the men together with the differences, the sigmas of the differences, and the quotient of the differences divided by the sigmas of the differences.

While none of the differences is statistically reliable, three of the differences do appear to be significant. The chances are about 99 in 100 that men will indicate as their best friends men who have more pleasing personalities than themselves. The chances are also 99 in 100 that men will indicate as their best friends men who are better adjusted socially than they are. And the chances are 93 in 100 that the best friends indicated are of a higher social intelligence than their own.

TABLE 2

SHOWING THE AVERAGES FOR THE MEN AND THEIR BEST FRIENDS, WITH THE DIFFERENCES BETWEEN THE AVERAGES, ETC.

Trait	Men	Friends	Diff.	Sigma of Diff.	Diff. Sigma Diff.
Pleasingness	63.94	68.63	4.69	1.89	2.48
Steadiness	52.00	50.81	—1.19	1.71	.70
Expressiveness	60.35	62.01	1.66	1.67	.99
Adjustment	61.81	66.81	5.00	2.17	2.30
Laird test	18.23	17.21	—1.02	1.12	.91
Army Alpha	145.94	147.56	1.62	3.96	.41
Thorndike	59.17	60.60	1.43	6.07	.24
Sims scale	27.96	27.00	— .96	1.62	.59
Social Intelligence	103.89	110.28	6.39	4.24	1.51

In securing the name of the best friend the subjects were merely asked to indicate whom they considered as their best friend. Whether the individual indicated was actually the best friend, of the length of the friendship, of the reciprocity of the friendly feeling, nothing can be determined. It is quite probable that college men indicate as their best friends persons whom they admire, and often persons whom they would like to emulate.

The evidence in Table 2, then, indicates specifically that college men strive to associate with and choose as their friends, or tend so to do, individuals who have more pleasing personalities than their own, who are socially better adjusted and who have more social intelligence than they themselves have.

If the popular belief is true that the friends one makes in college are more important than the facts he temporarily learns, this study indicates

TABLE 3

SHOWING THE AVERAGES FOR THE WOMEN AND THEIR BEST FRIENDS, WITH THE DIFFERENCES BETWEEN THE AVERAGES, ETC.

Trait	Women	Friends	Diff.	Sigma of Diff.	Diff. Sigma Diff.
Pleasingness	66.27	65.19	—1.08	1.42	.76
Steadiness	48.96	48.52	— .44	1.27	.35
Expressiveness	60.37	60.81	.44	1.70	.26
Adjustment	66.47	67.16	.69	1.71	.40
Laird test	17.65	18.09	.44	.71	.62
Army Alpha	149.41	147.58	—1.83	3.48	.53
Thorndike	60.61	60.39	— .22	2.97	.07
Sims scale	26.56	26.82	.26	1.15	.23
Social Intelligence	100.03	100.68	.65	8.07	.08

what the young man is striving for in his friendships. And we might ask the question, by the way: Can the college do anything to aid the college man to improve himself with respect to the pleasingness of his personality, his social adjustment, and his social intelligence?

Table 3 shows similar data for the women.

Here there are no reliable differences, nor any near reliable differences of any significance. It would seem, then, that the men in college seek opportunities to improve their personalities, to widen their adjustments, and improve their social information, while the women in college seem to be more satisfied with themselves in these respects. It may be argued, perhaps, that the women do not need the improvement in these respects that the men do, since a comparison of Tables 2 and 3 show that the women are more pleasing and better adjusted on the average. The differences between men and women in this respect, however, are not statistically reliable; although the chances are over 90 in 100 that the differences are reliable.² On the other hand, the average score on social intelligence for the men is a little higher than the average for the women.

Another fact that throws some light upon sex differences in friendships is that 38 of the 61 women showed reciprocal friendships, or about 62 per cent. Among the men, however, 18 of the 48, or about 38 per cent showed reciprocal friendships. Such a difference in the number of reciprocal friendships may indicate that women are less prone to give affection unless they receive it in return from the same individual; their friendships are perhaps more personal in nature, while the men show indications of greater hero worship than the women.

Combined with the fact that the friends indicated by the men have more pleasing personalities, are better adjusted, and have a higher social intelligence, the fewer reciprocal friendships may indicate that the men more than the women desire to get on by attaching themselves to persons who have more social influence than they have. Men get on by attaching themselves to and making use of powerful men; but women do not get on in the same way—they do not work through other women of greater social adaptability than themselves.

CONCLUSIONS

From this study we are justified in concluding that for both men and women birds of a feather are more likely to flock together than that opposites attract. This is particularly true with respect to pleasingness of personality, adjustment, and social intelligence.

Among the men, introverts tend to associate with introverts, and extroverts with extroverts.

²See (1).

Among the women, individuals with high Thorndike scores tend to associate with others with high Thorndike scores, and vice versa.

For sex differences we find that men tend to choose friends with more pleasing personalities than their own, who are better adjusted, and who have more social intelligence than themselves, whereas women do not.

There are more reciprocal friendships among women than among men.

REFERENCE

1. FLEMMING, E. G. Pleasing personality. *J. Soc. Psychol.*, 1932, **3**, 100-107.

29 Claremont Avenue
New York City

BOOKS

P. M. SYMONDS. *Diagnosing Personality and Conduct*. New York: Century, 1932. Pp. xvi+602.

When Goethe gave it as his opinion that personality is the supreme joy of the children of the earth, he could not have foreseen the joyless dissection of his romantic ideal one hundred years hence. There is certainly nothing poetical, metaphysical, or even humanistic in the modern psychological approaches to personality which are so conveniently surveyed in this volume. Without deference to philosophy, biography, or poetry, and with only a nod in the direction of psychoanalysis and the case study, Professor Symonds concentrates his attention upon the more "exact" methods of research, and describes each in all its quantitative glory. However little of the total field of personality he explores, his search so far as it goes is fruitful and his findings clear.

The reader must expect the book, in spite of its bulk, to be exceedingly limited in scope. Not only does it slight the non-quantitative methods, but it is markedly national in its outlook. There are more than nine hundred references in the bibliographies, but of these only 5% are to German publications and only 3% to publications in languages other than English and German. The German citations, furthermore, shrink to 2% if one omits the purely physiological references contained in the somewhat irrelevant chapter on measuring emotions. There can be no question that this narrow range of citations does an injustice to foreign work which, by and large, is not only prior to American work, but is likewise more sustained, more varied, and more profound. One looks in vain for mention of the diagnostic methods based on the researches of Jaensch, Rorschach, Bissky, Künkel, Huthe, Enke, Henning, Klages, Oseretzky, Spranger, Wolff, Bobertag, Aruheim, and other Europeans.

Granted then that the book is distinctly an American work, concerned almost exclusively with quantitative methods, it should be valued highly as a representative of this limited field of interest. Nowhere else can the student find so compact an account of the problems encountered in rating methods or in questionnaire methods. Other chapters contain serviceable surveys of free association, interviewing, the measurement of the environment, and like methods. The chapters on psychoanalysis and the case study, however, are meagre and unsympathetic. Weakest of all perhaps is the section on handwriting, contained in a chapter devoted to the external signs of conduct.

It is difficult to see how the author, after reviewing the work of Hull and Montgomery which bears only a remote relation to the claims and

methods of graphologists, can generalize so unguardedly. He writes, "the average of these correlations is $-.016$, which represents about the amount of assurance that one should give to the claims of graphologists" (527). He could have reported a number of studies contradictory in result and more adequate in method, to give a different and more discriminating statement of the case. Symond's attack on Saudek's methods of studying handwriting is quite unfair. It is made to appear that Saudek relies primarily on detailed "signs" in diagnosing script, whereas he is one of the most eloquent opponents of this primitive procedure. Continuing his attack on Saudek, the author argues that the well-known experiment in which Saudek picked successfully the script of dishonest employees, cannot be correct simply because it does not "square with other facts that are known concerning the specificity of conduct." Surely Saudek's experiment must be examined in its own right, quite apart from these "facts" of specificity which are decidedly tentative and quite possibly mistaken. It must be proved that Saudek's work is unreliable and that the other "facts" concerning specificity are both reliable and pertinent to the issue in hand.

From the theoretical point of view the crux of the book lies in its ardent defense of specificity. It is the author's opinion that, "from a review of all the skilful and ingenious methods for testing conduct directly that have been devised, the conclusion stands out above all others that conduct is very specific" (352). Although this question is too involved for a thorough treatment here, a few considerations, quite neglected by the author, must be urged on the other side.

1. The methods of current research upon which the doctrine of specificity rests are quite unsuited to the discovery of the fundamental consistencies within personality. Until adequate methods are devised for the study of the *complex* correspondences which lie within *complex* unities of personality, an exaggerated faith in specificity will prevail. What is forgotten is that segmental methods *can* yield only segmental results. Cleeton and Knight measure the length of noses but not the cast of countenance as a whole, Hull and Montgomery measure microscopically the width of pen-strokes but not the form-quality of the total script, Rich considers the hydrogen-ion concentration in saliva, but not the endocrine pattern or type, and Hartshorne and May deal with a psycho-socio-ethical *mélange* called "character" which gives little opportunity for the discovery of consistent *psychological* dispositions in personality.

If it is objected that one cannot *measure* pattern, form-quality, or style, the proper answer is, so much the worse for measurement. It is still in these complexities that one must seek congruence and integration, and any two variables arbitrarily isolated from the total structure for purposes of correlation will result only in an illusion of specificity.

2. As ill-adapted as most of the current methods are to the discovery

of consistency, they themselves yield copious evidence that the doctrine of specificity has been too hastily embraced. What else than personal consistency can the reliability of the scale signify? If a questionnaire contains a large number of items drawn from many representative fields of daily action, as most of them do, and if the subject answers these items consistently, which is what reliability means, where then is the "outstanding" specificity which Symonds sees in the results? There are scales whose split-half reliability and internal consistency reach $+.75$ or even $+.85$. These figures signify that most people respond to practically all the items in a consistent way. They are *characteristically* ascendant, extroverted, neurotic, or fair-minded. Their conduct is not specific.

Symonds himself recognizes the soundness of this interpretation. "If the instrument has high reliability, this is evidence that there is something there in reality, even if the name assigned to the test may not properly describe it" (22). Again, in discussing the A-S Reaction Study he writes, "the very reliability of the questionnaire indicates an internal consistency of the items which is significant..." (207). But in order to save the cause of specificity he suggests that the consistency may be due to a subjective urge of the people who take the scale to *appear* consistent. He recognizes the shortcoming of this interpretation, however, in the face of the validity of the scale.

The same type of reasoning applies wherever we find scales based on a wide variety of items and possessing a suitable internal consistency. They must be regarded as *prima facie* evidence for some kind of generality in conduct. The reliabilities of the better constructed scales as reviewed by Symonds, for example on p. 185, by no means support his statement on p. 352 that "a battery of tests designed to test such a trait as persistence, or aggressiveness, or speed of decision gives results so varying and with so little consistency as to furnish little warrant for assuming the presence of such a trait." The author himself shows that the internal reliability of several scales is very satisfactory, and therefore by his own word should see "something in reality," beyond mere specificity.

Occasionally, of course, clear-cut specificity is encountered. "Honesty" as studied by the C. E. I. is an example. When such results are found they may mean no more than that the scale was not adapted to the discovery of true psychological dispositions. But even when specificity is unequivocally established in one field, there is no ground for a sweeping denial of consistency in all other fields. *Occasional* specificity is not incompatible with the doctrine that generality prevails in the organization of personality.

3. The doctrine of specificity rests in large part upon research with children who must, of course, be expected to show little consistency, but the theory should not be applied without caution to adult conduct. Hartshorne

and May who employed children in the fifth to the eighth grades admit that even in this narrow range of age there is evidence for greater consistency in older children, as common sense would surmise. Yet Symonds prefers a less obvious interpretation than that of progressive integration with age; he writes that the "increase in reliability with increasing age may be explained as the growth of the ability of individuals to diagnose and judge their own states." Granted that such insight may develop with increasing age, is not this in itself evidence of the presence of a more highly evolved integration?

4. A further difficulty with the doctrine of specificity springs from its dependence on quantitative methods of *mass* investigation. Even if we suppose for a moment that every measurement of internal consistency of every scale were zero (and this is of course not the case), what would it prove? It would show only that the subjects were not uniform enough in their behavior to be considered comparable. It would not show that each individual subject is not consistent *in his own way*. Scalable traits are not the only kind of traits. There are all manner of intra-individual patterns of consistency, strictly personal traits, which evade the "personality test." The appearance of specificity in such a test would spring solely from the fact that men are different and not comparable, and not at all from the fact that the conduct of each individual man is itself totally unorganized and specific. Whatever the reliability of a scale, the phenomenon of unique, concrete individuality remains untouched and undiscovered by mass methods.

In the field of psychology only psychoanalysis and certain German schools have made even a tentative approach to the study of the unique harmonies of personality. Their methods, however imperfect, reveal little specificity and great consistency. Symonds frankly distrusts these "clinical" methods because of the "fallacy of drawing conclusions from the single case" (7); but his objection withers in the face of Lewin's demonstration that one can often draw valid conclusions for a single case from a study of that case *and no other*. Specificity can be assumed as the basic law of human personality only after specificity has been shown to hold for *individuals* as well as for masses. If it cannot be discovered by the direct study of persons, it is probable that specificity is after all an artifact resulting from the methods employed. Measurements of conduct *en masse* are so circuitous as to be almost irrelevant to the issue of the organization of personality.

5. Considering the indirectness and general ineptness of statistical methods, it seems almost surprising that the internal consistencies of scales are as high as Symonds reports them to be. Even the lower ones do not seem to justify his emphasis upon specificity. When an uncorrected reliability coefficient of, say, $+0.45$ is encountered, what does it signify?

Generality or specificity? At this point the objective method vanishes, and subjective interpretation sets in. If the impossible ideal of perfect unity is set for measures of reliability, this coefficient of course appears low, and it is said that the consistency of behavior represented by the scale is negligible. But if the investigator has in mind all of the inadequacies of the statistical method in uncovering the true structure of personality, and all the errors of measurement in his clumsy devices, such a coefficient will indicate to him not specificity but a promising degree of generality.

6. If we become intentionally naïve and ask whether any specificist in daily life can believe and be guided by his theory, we discover swiftly that he cannot. Logically a specificist should never allow himself to use an adjective to describe a person, but only to describe a single act at a certain time. What adherent of the doctrine is so steadfast in his faith and practice? There are many places in the book where the author speaks of traits and consistency sympathetically, holding for example that "certain traits may be rated more reliably than others" (106), and accepting Shen's finding that the "tendency to overestimate or underestimate the self is more or less consistent with the individual" (109). In another place he regards interest questionnaires as promising because they point out "certain broad trends of interests" (256). But as conclusive a self-contradiction as any, quite typical of the difficulties which a specificist makes for himself, is Symonds' own statement that individuals can be trusted fairly well not to falsify their answers to a questionnaire, for "*the truthful person finds it very difficult to be untruthful*" (italics mine) (122).

7. In spite of the inconsistencies just pointed out, Symonds is an implacable foe of the concept of "trait." Just what is he opposing? First of all he believes "characterization is rough and crude when it is accomplished by means of descriptive adjectives," and that "even so remarkable a character as Lindbergh probably does not always show deliberation or industry to the same degree at all times or in all situations" (50). The most enthusiastic "trait psychologist" could endorse these views, but at the same time he would see some justification, as Symonds does not, for the economisms of Lindbergh's "purposefulness," "seriousness," "deliberateness," "stability," "efficiency," and "industry." It is doubtful whether a hero who is a bundle of unrelated and specific habits depending for their functioning upon the situation of the moment, would have been able to go through such a variety of ordeals *all new in type to him* (flying the ocean, unprecedented ovations, kidnapping, and newspaper persecution), all the while displaying consistent levelheadedness, modesty, and poise.

No "trait psychologist" believes that *every* act is always in perfect correlation with every other. If it were, every personality could be given a single "score" and dismissed. The stimulus of the moment must be allowed for, as well as dissociations among habits and inequality in abilities. The

most assertive person may become submissive and silent in the presence of his physician. But occasional lapses do not affect the evidence for a general level of adjustment which is implied in the concept of trait. They simply increase the complexity of the problem of the organization of personality.

The "hypostatization" of traits and the "traitification" of behavior are to Symonds' mind great menaces. Traits he considers as "solidifications" of our own imagination; they are in all respects as treacherous as the faculties of old. Now, the following considerations should lay the facultative ghost.

a. In the first place, traits, as conceived at the present time, are meaningful personal dispositions, empirically discovered, and are not arbitrary constructs as were the older faculties. There is a world of difference between ascendance, expansiveness, extroversion, and Sensation, Will, and Reason. Traits are thought of as rooted in the familiar processes of neural integration; faculties depend upon philosophical and theological decisions as to the "ultimate" compartments of mind.

b. Faculties were usually considered to be quite independent of one another. No psychologist with an understanding of the complexities of neural patterning would hold any such view regarding traits.

c. It should also be borne in mind that the charge of "verbal fiction" may be applied to faculties but not to traits. Traits are dynamic dispositions discovered by experimental and by clinical research. The fact that names for these dispositions are necessary before they can be made available for scientific discourse is a misfortune, not peculiar to them, but shared as well by every natural object or event to which science turns its attention. Of course a trait-name does not guarantee that a corresponding trait exists in every person at all times. Whoever held that it did?

d. If the specificist still cries "faculty" let him examine his own position. He sees personality as an unordered aggregate of countless stimulus response units, insulated, specific, and not generalized. Are these minute elements less facultative than traits? Does a small illegitimate baby represent a lesser sin than a large one?

In answer to many of these arguments the author would undoubtedly refer to his earlier, more theoretical work, *The Nature of Conduct*, in which he attempts to account for some of the indisputable evidence for generality while holding tenaciously to the theory of specificity. He contrives to do this with the aid of his doctrine of 'confacts.' A 'confact' is the speculative offspring of the theory of specificity and the theory of identical elements. I confess to my inability to see the advantage of the concept of confact over the concept of trait. The inappropriateness of the parent doctrine of identical elements has been discussed elsewhere (G. W. Allport and P. E. Vernon, *Studies in Expressive Movement*, Chap. VII.).

Enough has been said in disagreement with Symonds' theory of per-

sonality. As a practical reference book, summarizing in a convenient and timely fashion the mountains of current "exact" research in personality, the volume has an important place. I have found it useful as a textbook in courses in the psychology of personality. (It is, however, somewhat distressing to find that students who read the book receive a vigorous conviction that "reliability" is of basic importance in investigations of personality, without at the same time obtaining a clear understanding of the meaning of reliability. Unless I am mistaken, Professor Symonds has neglected to define the most important concept in his book.)

At the risk of seeming captious to an extreme degree, I beg to venture a final, and perhaps merely personal, expression of dissatisfaction with the author's desire to apply the present results of research in personality to education, business, and criminology. "We are interested," he writes, "in selecting persons who will be successful in business and industry, or in individuals so that they may make the most happy occupational adjustments . . ." (pp. 300 f.). Ratings repeated systematically in industry, we are told, "keep the personnel spirit alive. . . ." It seems to be a mistake at this early stage to confuse research in personality with the ulterior interests of character education, social reform, and industrial exploitation. Particularly it seems to me that it is exceedingly harmful to give advice to educators or to industrialists based on the tentative and insecure hypothesis of specificity. Psychological doctrines which are only a few years old are notoriously subject to error. It is not a service to society to apply theories which are tentative and so clearly destined to revision or to total rejection.

GORDON W. ALLPORT

Harvard University
Cambridge, Massachusetts

BOOKS RECEIVED

- BEHRENDT, R. *Politischer Aktivismus: ein Versuch zur Soziologie und Psychologie der Politik*. Leipzig: Hirschfeld, 1932. S. vi+182. RM. 5.80.
- BINGHAM, W. V. [Ed.] *Psychology today: lectures and study manual*. Chicago: Univ. Chicago Press, 1932. Pp. xii+272+44+48+32+32+32+36. \$1.50.
- BOSSARD, J. H. S. [Ed.] *Man and his world*. New York & London: Harper, 1932. Pp. xiv+755. \$3.50.
- BURROW, T. *The structure of insanity*. (*Psyche Miniatures, Gen. Ser.*, No. 40.) London: Kegan Paul, 1932. Pp. 80.
- COWLEY, W. H. *The personnel bibliographical index*. Columbus, Ohio: Bur. Educ. Res., Ohio Univ., 1932. Pp. 433. \$4.00.

- DICKINSON, R. L., & BEAM, L. A thousand marriages: a medical study of sex adjustment. Baltimore, Md.: Williams & Wilkins, 1931. Pp. xxvi+482. \$5.00.
- EXNER, M. J. The sexual side of marriage. New York: Norton, 1932. Pp. 252. \$2.50.
- FRANK, G. Thunder and dawn: America's appointment with destiny. New York: Macmillan, 1932. Pp. xiv+404. \$3.50.
- GAULT, R. H. Criminology. Boston: Heath, 1932. Pp. x+461. \$3.48.
- HERSEY, R. B. Workers' emotions in shop and home. Philadelphia: Univ. Pa. Press, 1932. Pp. xviii+441. \$3.00.
- JASTROW, J. The house that Freud built. New York: Greenberg, 1932. Pp. x+293.
- LANDMAN, J. H. Human sterilization. New York: Macmillan, 1932. Pp. xviii+341. \$4.00.
- MURCHISON, C. [Ed.] A history of psychology in autobiography: Vol. 2, by Bourdon, B., Drever, J., Dunlap, K., Ferrari, G. C., Franz, S. I., Groos, K., Heymans, G., Höffding, H., Judd, C. H., Morgan, C. L., Pillsbury, W. B., Terman, L. M., Washburn, M. F., Woodworth, R. S., Yerkes, R. M. Worcester, Mass.: Clark Univ. Press; London: Oxford Univ. Press, 1932. Pp. xx+407. \$5.00.
- NEPRASH, J. A. Brookhart campaigns in Iowa: 1920-1926. A study in the motivation of political attitudes. (*Stud. Hist., Econ., & Pub. Law.*, No. 366.) New York: Columbia Univ. Press; London: King, 1932. Pp. 128. \$2.25.
- PILLSBURY, W. B. An elementary psychology of the abnormal. New York & London: McGraw-Hill, 1932. Pp. x+375. \$2.50.
- SELIGMAN, E. R. A., & JOHNSON, A. [Eds.] Encyclopaedia of the social sciences. Vol. 4: Com-Dan. New York: Macmillan, 1931. Pp. xxviii+710. \$7.50.
- SELIGMAN, E. R. A., & JOHNSON, A. [Eds.] Encyclopaedia of the social sciences. Vol. 5: Dan-Exi. New York: Macmillan, 1931. Pp. xxiv+690. \$7.50.
- SELIGMAN, E. R. A., & JOHNSON, A. [Eds.] Encyclopaedia of the social sciences. Vol. 6: Exp-Gos. New York: Macmillan, 1931. Pp. xxv+713. \$7.50.
- SELIGMAN, E. R. A., & JOHNSON, A. [Eds.] Encyclopaedia of the social sciences. Vol. 7: Gas-Ind. New York: Macmillan, 1932. Pp. xxvi+722. \$7.50.
- STONE, C. P., DARROW, C. W., LANDIS, C., & HEATH, L. L. Studies in the dynamics of behavior, ed. by K. S. Lashley. (*Behav. Res. Fund Monog.*) Chicago: Univ. Chicago Press, 1932. Pp. xiv+332. \$5.00.
- WARDEN, C. J. The evolution of human behavior. New York: Macmillan, 1932. Pp. xii+248. \$3.00.
- WHEELER, R. H., & PERKINS, F. T. Principles of mental development: a textbook in educational psychology. New York: Crowell, 1932. Pp. xxviii+529. \$3.75.
- YATES, D. H. Psychological racketeers. Boston: Badger, 1932. Pp. 232. \$2.00.
- YOUNG, P. V. The pilgrims of Russian town. Chicago: Univ. Chicago Press, 1932. Pp. xx+296. \$3.00.

SOME PROPERTIES OF THE THURSTONE PERSON- ALITY SCHEDULE AND A SUGGESTED REVISION*

RAYMOND ROYCE WILLOUGHBY

The relatively high reliability and the wide range of scores yielded by Thurstone's (7) neurotic inventory have stimulated a substantial amount of work with it, and its characteristics as a complete scale are as a result fairly well known. The characteristics of the individual items and of their various possible groupings, however, are not so well known, and a certain practical interest attaches to this aspect of the matter by reason of the fact that the entire scale of 223 items is uncomfortably long (from the standpoint of the subject) and difficult to associate with other items in a comprehensive testing program, so that an abridgment is desirable. Further, the objection has been offered that the scale is an "omnibus" measure, and that no one knows whether the "trait" measured is even approximately unitary. Both these questions would be somewhat clarified by an analysis of the incidence characteristics of the separate items, their discriminatory power, and the tendency to association between them; and the publication of certain recent work on these matters, together with some original researches in connection with a larger investigation, appear to make a systematic consideration of them timely.

PREVIOUS WORK; WITH AN EVALUATION OF INCIDENCE AS A CRITERION OF VALIDITY

So far as a somewhat cursory examination of the literature reveals, the original framers of the items included in the inventory submitted no quantitative data on them specifically. Thurstone's original article, however, includes data from his Chicago freshman population based on discriminatory power; this is measured by the difference between the incidence percentages from the highest and lowest total scoring sub-populations; i.e., if A is the sub-population making the lowest total scores (say the lowest 50 in a population of 700), E the sub-population making the highest total scores, p the proportion of maladjusted answers on a given item in E , and q the corresponding proportion in A , then $p-q$ is the measure of discriminatory power

*Accepted for publication by Carl Murchison of the Editorial Board.

for that item. The measure is thus a sort of correlation between the item and the total score, or an indication, somewhat indirect, of its representativeness. On this basis Thurstone selects 42 items as being most discriminatory. Harvey (4) finds this measure necessary, but insufficient; he suggests that the absolute size of the incidence percentage must be taken into account, and that the larger it is the more desirable the item. He offers no data in support of this judgment, which seems reasonable enough within limits.¹

It may be tested, however, by a device which we take from Mrs. Thurstone (9), viz., by correlating scores for sections of the incidence range against the total score. This will of course give coefficients that are partly "spurious" (in the sense that we know their source), since the section correlated appears also in the whole; but since the discovery of a section of the range highly correlated with the whole would make possible its substitution therefor, this "spurious" correlation is exactly what is desired, i.e., a measure of the extent to which the whole can be replaced by the section.

The frequency distribution of item incidences for a population of 144 freshman women is shown in Table 1.

TABLE 1

%0 ³	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	14	5	16	11	19	19	10	15	17	7	13	7	7	6	8	6	3
36	38	40	42	44	46	48	50	52	54	56	58	60	62	64	66	Total	
3	6	5	2	2	5	3	2		1		1	2			1	223	

²Index is lower limit of class throughout the paper.

Thus the incidences of about half the items fall below 19%, and those of about three-fourths below 28%. The impression left by this state of affairs is not favorable, since, with the median item occurring in only one subject out of five, it appears unlikely that the diagnostic power of most of the items can be considerable. Harvey regards this as self-evident: "... a measure of the degree to which

¹It seems, however, to imply an assumption that statistical convenience is superior as a criterion to clinical facts, which is hazardous as a general conclusion. Were the feeling of being watched on the street found to have as low an incidence as 20% and the enjoyment of puzzles as high an incidence as 40%, it is likely that the former should still be preferred to the latter as a diagnostic item; naturally this consideration does not apply in the very low ranges, as in items with incidences of 1 or 2%.

the item itself prevails as a psychoneurotic symptom in a relatively unselected population. Obviously, if low, it cannot be a very characteristic neurotic symptom."

We shall sample the range at five points, separated by intervals of 4% ; four of the sections will include 6%, the fifth, due to special circumstances, about 28%. The full description of the sections is presented in Table 2, together with the correlations with the whole scale.

TABLE 2

Incidence range	Items ^a	<i>r</i>
0-5%	27, 56, 60, 67, 69, 75, 87, 101, 102, 103, 104, 107, 116, 132, 133, 134, 135, 160, 165, 198, 211, 212, 215, 216, 217, 218, 219—27 items.	.59 ± .04
10-15%	5, 9, 11, 14, 15, 17, 26, 29, 34, 36, 37, 40, 41, 43, 45, 53, 55, 59, 61, 63, 64, 69, 71, 83, 84, 85, 97, 106, 111, 119,	.87 ± .01
20-25%	1, 2, 18, 23, 30, 39, 42, 44, 46, 50, 95, 108, 110, 114, 115, 126, 129, 137, 144, 146, 147, 149, 155, 170, 179, 185, 189, 191, 193, 200, 204, 210, 213—48 items. 120, 127, 153, 156, 163, 164, 171, 178, 182, 183, 194, 207—27 items.	.86 ± .01
30-35%	13, 19, 20, 57, 79, 85, 91, 100, 121, 159, 161, 175, 180, 190, 197, 220, 221—17 items.	.79 ± .02
40-67%	3, 16, 21, 24, 31, 49, 70, 73, 80, 89, 90, 92, 94, 138, 145, 148, 150, 151, 152, 177, 181, 187, 195, 206—24 items.	.81 ± .02

^aNumbered straight through the scale.

It thus appears, contrary to expectations, that the representativeness of a group of items is not greatly affected by their incidences; even in the lowest range, where the coefficient is sensibly lower than for the others, this effect seems to be partly due to the bunching of cases at zero, an effect which might have been prevented by taking a larger number of items. The best validities occur in the neighborhood of the median incidence, but all the coefficients are good, and there seems no reason why satisfactory sub-scales may not be constructed using items from anywhere above the lowest incidence levels. A theoretical generalization of possible importance is that any given neurotic symptom occurs with relative infrequency, or at least is recognized by the subjects with relative infrequency, so that in the construction of a two-response scale for the detection of neu-

roticism it will ordinarily be necessary to assemble a list of symptoms considerably longer than the frequency of modal response desired. This is a partial justification for the great length of the Thurstone scale, but it is likely that the difficulty may be avoided by using a multiple-response device. Another generalized consequence is that the trait measured by the scale ("neuroticism") exhibits, so far as it can be measured, more extreme and less frequent deviations in the maladjusted than in the adjusted direction, and that this is not, as might be hypothecated, a consequence of the downward limitation of the scores by zero, but a real result of the relative infrequency of the symptoms or of their recognition.

Harvey used his combined criteria of incidence and discriminatory power with a population of 146 University of Texas sophomores (male); he found 30 items (barring duplicates) of special value. He also divides the items (logically) into a rather large number of categories (11), which overlap somewhat, and shows that the scale is unevenly weighted with respect to these categories; for this condition, which he regards as a defect, he suggests no specific remedy. In this judgment he appears to neglect the possibility that the neurotic constellation, which it is desired to measure, may really be weighted unevenly, i.e., may be manifested more often and more markedly in some fields of behavior than others; so that the important consideration here is the manner in which items distribute themselves in terms of the responses of subjects rather than in terms of logical categories.

The recent work of Darrow (3) is especially important. Basically it is an attempt to associate physiological syndromes with constellations of items from the inventory; and while the coefficients are much too low for predictive purposes, we must, I think, regard the attempt as a success inasmuch as Darrow obtained clear-cut correlation between certain syndromes and constellations, and equally clear-cut absence of correlation between these syndromes and other constellations. That the attempt involved a good deal of "cutting and trying" is not evidence, as might be at first thought, of artificialism in the results, but a consequence of the endeavor to approximate whatever natural groupings exist in the material. After assembling his basic data, Darrow proceeds by examining the items which are related to given physiological measures, and infers that two items related to the same measure, or two measures related to the same item, are probably significantly associated. The relationships of

other items and measures to this nucleus are then studied, and in this way constellations of items and syndromes of measures are built up. The empirical constellations of this sort in which we shall be most interested, since we are concerned with neuroticism, are those denominated by him the socially inactive, the neurasthenic, the hyper-sensitivity, the depression, and the anxiety personality constellations; there are five others which would be supposed a priori to have a low relation to each other and to the rest of the scale, and this supposition is confirmed by the results in all but one case. Synthesis of these constellations into larger ones on the basis of correlations with the syndromes yields the neurotic constellation (composed of the five primary constellations mentioned above) and the melancholic constellation (depression, anxiety, hyper-sensitivity). The correlation between neurasthenic constellation and neurasthenic syndrome is .38; the neurotic syndrome correlates with melancholic constellation .37, with neurotic constellation .36, with hyper-sensitivity constellation .35, with anxiety constellation .34, with neurasthenic constellation .33, with socially inactive constellation .30, with depression constellation .30, with total neurotic score .30; and not at all with either intelligence or extroversion. These correlations (.38 to .30) are higher by a comfortable margin than any others reported in the study, and their general correspondence with the highest physique-personality correlations previously reported suggests that Darrow has succeeded in wresting from the data their highest correspondences. The particular importance of the study from the present point of view is that a purely empirical method, based on physiological approaches, yields definite and therefore presumably natural groupings of items.

The approach of Bernreuter (1), who determines four different (?) "traits" by differential weightings of the same responses, does not seem promising, but must await publication of his data for evaluation.

The writer has used the schedule on a population of married couples ($N = 152$) and, since no female student control population had been reported in such form that item-by-item analysis was feasible, he has constructed one from the records of 144 unselected University of Pittsburgh freshman women of the class entering in 1930, kindly furnished by Dr. Norman L. Munn. There are thus available adequate incidence data on three new populations of about 150 individuals each. In addition, the cohesiveness of the scale has been

studied by intercorrelating for the husbands and wives the partial neurotic scores of six logically derived sub-scales, exhaustive and exclusive, and subjecting them to the Spearman two-factor and the Thurstone multiple-factor analyses; also by a study (for the Pittsburgh women) of the association between items, from which has emerged an empirical constellation of closely related items. The latter results are incomplete, inasmuch as the 40 items of highest incidence were selected for the study, in order to minimize the labor involved in examining all possible associations among the 223 items of the total scale.

SUB-SCALES; REPRESENTATIVENESS, INTERCORRELATION, AND FACTOR ANALYSIS

The items included in each of the six categories mentioned above are as follows:

Social (So); items dealing with subjects' reactions to human environment, as 57 *Are you troubled with shyness?*

1, 2, 3, 10, 12, 13, 15, 17, 18, 25, 28, 30, 33, 36, 41, 43, 44, 57, 61, 67, 70, 72, 80, 83, 89, 91, 94, 98, 99, 100, 112, 115, 116, 119, 120, 123, 126, 131, 138, 140, 143, 145, 150, 157, 161, 169, 173, 175, 178, 180, 181, 182, 187, 195, 196, 205, 207, 209, 215.

Extrovert (Ex); items dealing with subjects' reactions to non-human environment, as 167 *Are you systematic in caring for your personal property?*

14, 20, 32, 38, 46, 54, 59, 96, 118, 121, 146, 167, 170, 183, 190, 192, 200.

Fantasy (Fa); items dealing with subjects' inner experience, as 39 *Are you frequently burdened with a sense of remorse?*

6, 11, 16, 19, 21, 23, 24, 26, 31, 35, 39, 40, 47, 51, 52, 53, 55, 58, 63, 64, 65, 71, 73, 74, 75, 76, 79, 84, 85, 86, 92, 97, 101, 103, 104, 105, 106, 108, 111, 113, 114, 124, 127, 129, 136, 139, 141, 142, 148, 149, 151, 153, 156, 158, 164, 168, 174, 177, 184, 185, 186, 193, 194, 197, 199, 201, 202, 206, 211, 214, 219, 220.

Physical (Ph); items dealing with subjects' somatic phenomena, as 198 *Did you ever have anemia badly?*

5, 7, 9, 27, 34, 37, 42, 48, 49, 50, 56, 60, 62, 66, 78, 82, 87, 88, 90, 93, 109, 110, 117, 132, 133, 135, 144, 152, 154, 155, 160, 162, 163, 166, 176, 179, 189, 191, 198, 203, 204, 208, 210, 213, 216, 218, 223.

*Parental*⁴ (Pa); items dealing with the subjects' immediate family, as 188 *Was your mother the dominant member of the family?*

4, 8, 22, 45, 69, 77, 81, 102, 107, 122, 134, 147, 159, 165, 188, 212, 217, 221.

⁴These categories are as nearly as possible identical with the only two set up by Thurstone in his original article.

*Sex*⁴ (Sx); items dealing with subjects' sex attitudes, as 68 *Have you ever been afraid that you are sexually inferior to other men (other women)?*

29, 68, 95, 125, 128, 130, 137, 171, 172, 222.

The objective in this classification was to obtain a small number of more or less cohesive constellations; the data on intercorrelations given below (page 408) show that they are in fact considerably more cohesive than chance alone would determine. The correlations of each category with the total neurotic score are given in the following table (Table 3), and in connection with the preceding material on incidence *vs.* representativeness may shed some additional light on the problem of separating out representative sub-scales.

TABLE 3

	H	W
Fa	.90 \pm .01	.87 \pm .01
So	.79 \pm .02	.76 \pm .02
Ph	.72 \pm .03	.70 \pm .03
Pa	.52 \pm .04	.40 \pm .05
Ex	.64 \pm .03	.47 \pm .04
Sx	.58 \pm .04	.38 \pm .05

The categories So and Fa thus correlate by far the best with the entire scale, and this seems to support our criticism of Harvey's objection, mentioned above, that the original scale is unequally weighted. It appears, in fact, that "neuroticism" is chiefly a peculiar kind of maladjustment in the fantasy and social spheres, with some hypochondriacal aspects. The remaining coefficients (Pa, Ex, Sx), especially in view of the "spurious" element present in all of them, are of negligible size.

The cohesiveness of the scale may be measured (1) by the intercorrelations of its parts and the types of factor analysis of these, or (2) by the degree of association between specific items. The intercorrelations (for husbands and wives) will be presented first, after which analysis will be undertaken by the methods due to Thurstone (8) and to Spearman (as interpreted by Holzinger, 5); later a partial analysis of the association of items, following that of Beyle (2), will be attempted for the Pittsburgh women.

The categories are intercorrelated as follows (Table 4), the

TABLE 4

	Ph	So	Fa	Pa	Ex	Sx	
Ph		.34	.59	.36	.46	.46	
So	.38		.57	.23	.40	.47	(N = 152)
Fa	.59	.55		.43	.60	.44	C ₁ C ₂
Pa	.08	.22	.22		.34	.25	C ₁ .67
Ex	.37	.39	.40	.19		.33	C ₁ .71
Sx	.08	.32	.20	.17	.29		

values for the husbands being presented in the upper part of the table and those for the wives in the lower.

Control categories have been formed consisting of groups of 46 and 17 items (C_1 and C_2 respectively) selected (without duplication) by a random method from the entire 223; the item numbers included are:

C_1 : 1, 2, 7, 9, 10, 13, 17, 21, 22, 29, 32, 45, 52, 53, 55, 62, 78, 80, 88, 89, 97, 99, 104, 122, 126, 132, 136, 137, 141, 143, 145, 149, 153, 155, 159, 165, 170, 176, 178, 182, 189, 205, 207, 216, 222, 223.

C_2 : 6, 35, 74, 81, 112, 115, 131, 146, 150, 152, 156, 163, 183, 186, 210, 212, 221.

The correlation between these random categories, it will be observed, is significantly higher [$d/PE_d = 4.6$ (husbands), 7.3 (wives)⁵] than those between the logically erected ones; this is clear evidence that the latter represent constellations of items which are more nearly independent psychologically (though of subjective and logical origin) than chance would determine, although they are still strongly associated (as the several lines of evidence presented indicate that virtually all parts of the scale are).

The multiple-factor analysis of Thurstone is designed to yield an indication of the number of general uncorrelated factors which it would be necessary to postulate in order to account for an observed set of intercorrelations, and to secure an expression of the proportion of each ("loading") present in each test (here, category). Thurstone presents a simple method for determining in advance of the calculations the approximate number of general factors necessary; the application of this method to the present data results in an esti-

⁵ d_r was computed by subtracting from r_{0102} the mean of the category intercorrelations (husbands .42, wives .30); the PE of this mean r was obtained by the usual formula, as though it were an original datum.

mate of three general factors for the husbands and four for the wives. Calculation yields the first factor loadings presented in Table 5.

TABLE 5

	Ph	So	Fa	Pa	Ex	Sx	M
H	.77	.70	.84	.63	.72	.68	.72
W	.65	.74	.77	.49	.68	.53	.64
M	.71	.72	.81	.56	.70	.60	.68

Table 6 shows the squares of these, which according to Thurstone's analysis should be measures of the extent to which the categories are accounted for by this first general factor.

TABLE 6

	Ph	So	Fa	Pa	Ex	Sx	M
H	.59	.49	.71	.40	.52	.46	.52
W	.42	.55	.59	.24	.46	.28	.41
M	.50	.52	.66	.31	.49	.36	.46

The selection of sub-group s (Thurstone's step 3), defined as all the tests that correlate positively with test a (here Fa), includes the entire series. But sub-group t (step 7), all tests that correlate positively with test b (here Ex), also comprises the entire series; Σt_1 , the sum of the projections on the axis representing the first factor (i.e., sum of the first factor loadings) of the tests in sub-group t , is therefore equal to Σs_1 , the sum of the first factor loadings for all the tests; and Σr_{tt} , the algebraic sum of all the coefficients in a full table of the tests in group t , is identical with Σr_{ss} , the same function for the tests in group s . Since $\Sigma s_1 = \Sigma r_{ss}$ (Thurstone's formula 11), it follows that Σt_2 vanishes, making all the second factor loadings indeterminate. The experiment was tried of limiting the composition of sub-groups s and t to those tests correlating above .25 with the pivot tests (using wives only); the first factor loadings are here (Table 7) compared with those derived by the published method.

The difference is considerable only in those tests which in the fuller determination do, and in the more limited do not, have an influence in the determination of the factor axis. Sub-group t , however, re-

TABLE 7

	Ph	So	Fa	Pa	Ex	Sx	<i>M</i>
above .00	.65	.74	.77	.49	.68	.53	.64
above .25	.76	.76	.83	.23	.71	.29	.60

mains identical with sub-group *s*, and accordingly the modification has not succeeded in removing the indeterminacy from the second factor loadings. It is concluded, therefore, that the present situation of a small group of tests positively and rather highly intercorrelated constitutes a special case unfitted for treatment by the methods so far announced by Thurstone, and probably among those to which he makes reference as to be treated later (p. 425). The attempt, however, yields evidence that the scale scores may be thought of as accountable to the extent of about half by a single general factor; this factor is especially strong in the fantasy category, and relatively weak in the categories of sex and parent items, whence the most parsimonious hypothesis appears to be that the factor is an inner anxiety or hyper-sensitivity, a finding much like that deducible from Darrow's work, summarized above. It is of interest that although the neurotic scores of the wives are significantly higher than those of the husbands, the proportion of these scores accountable for in terms of the postulated inner anxiety is somewhat less for them than for the husbands, suggesting possibly an influence of habit superposed upon psychological tendency; of special interest is the reversal of the saturation with the general factor in the categories of physical and social items, as between the sexes; it is possible that the husbands' tendency to anxiety is expressed in physical, the wives' in social maladjustments; the greater importance of the latter for women has been shown in other studies. General confirmation of these results from the Thurstone analysis will be found in the more limited analysis of Spearman; in the meantime, a graph (Figure 1) is presented for the saturations (squares of the loadings) of the categories with the general factor here found.

The two-factor analysis, as is well known, is more limited than Thurstone's, being confined to the pattern of a general factor running through the variables plus a factor specific to each; the criterion for the acceptability of this pattern is that the distribution of tetrad differences (of which $t_{1234} = r_{12}r_{34} - r_{13}r_{24}$ is the type) shall be completely ascribable to random sampling, i.e., the mean tetrad shall be insignificant with respect to its probable error.

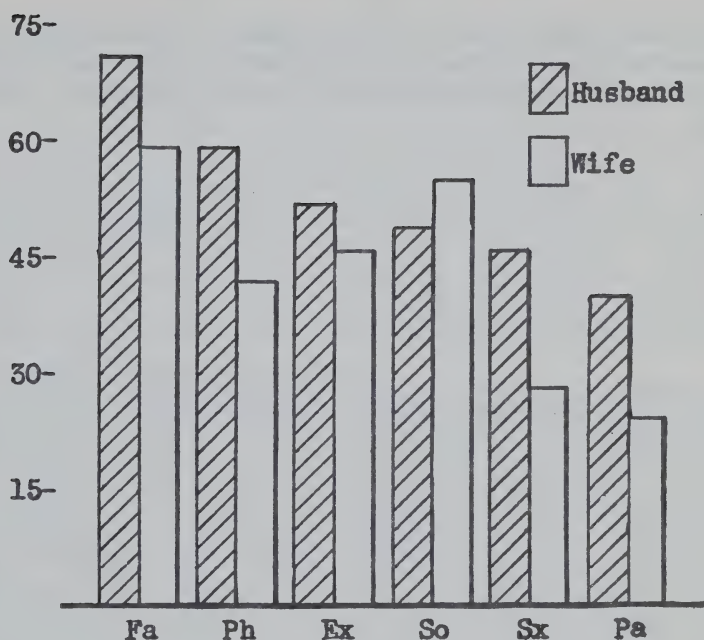


FIGURE 1

Squares of the loadings of each category with the first general factor (Thurstone analysis); i.e., proportion of the variance (?) of each which may be thought of as due to this factor.

In the present case the distribution of tetrads for the husbands has a mean of $-.041$ and a standard deviation of $.068$, while that for the wives has a mean of $.026$ and a standard deviation of $.048$. The average probable errors of the set of 45 tetrads are $.028$ for the husbands and $.026$ for the wives. The acceptability of the two-factor pattern is therefore established. It is next desirable to determine the "saturation" (corresponding to the "loading" of Thurstone) of each variable with the general factor; this is given by Holzinger's (5) formula 8, which when applied to these data yields the following saturations (Table 8):

TABLE 8

	H	W		H	W
Ph	.69	.52	Pa	.46	.29
So	.60	.72	Ex	.65	.61
Fa	.86	.76	Sx	.59	.33

The best weighted pool of the six categories yields multiple correlations of .92 for the husbands and .89 for the wives with the general factor (Holzinger's formula 12); and the appropriate weights are:

TABLE 9

	H	W		H	W
Ph	1.3	.7	Pa	.6	.3
So	.9	1.5	Ex	1.1	1.0
Fa	3.3	1.8	Sx	.9	.4

In general, the results of the Spearman analysis clearly confirm those by the Thurstone method in showing a marked degree of cohesion and a strong indication of the presence of at least one general factor. All methods of approach also demonstrate clearly the predominance of the fantasy category in determining the scale as a whole. The social and physical categories appear to be next most important, with the extrovert, sex, and parental categories of distinctly minor consequence. It will be recalled in this connection that Thurstone's original statement, to the effect that the neurotic personality is one which fails in some manner to express itself upon external reality and so returns upon itself, was confirmed by Harvey, and that Darrow's analysis also led to the postulation of socially inadequate, depressed, anxious, and somasthenic constellations, all tending to independence and significance. In short, all lines of attack lead to a point from which the general outlines of a clinically and statistically unitary trait begin to emerge; and this trait, as was well stated by Thurstone in his original article, is most accurately represented by the items which we have denominated fantasy and (rather unexpectedly) least so by those we have designated parental. Still another line of approach will be found to confirm this result.

COHESIVENESS

The ingenious but somewhat cumbersome methods used by Beyle (2) were devised by him (evidently without knowledge of other similar approaches) as a tool with which to determine the existence and basis of significant associations ("blocs") among legislators voting on different issues. They rest ultimately upon the same basis as does the coefficient of contingency, viz., the excess of observed

over chance frequency cell by cell in a two-dimensional association table. In our judgment, Beyle implied for them a somewhat greater power than can be demonstrated, since the phenomena under examination prove sooner or later to constitute a continuum rather than a collection of groups, and the division into "central nucleus of principal bloc," "inner and outer fringes," "secondary blocs," etc., must in general be made somewhat arbitrarily, as by the drawing of a line where the first blank occurs, etc.; nevertheless they do succeed fairly well in ranking the individuals in order of their cohesion. The present adaptation is somewhat free and the analysis based upon it only partial, in view of the prohibitive amount of labor involved in a complete examination of the data.

The basic determination is that of the chance association between two items. It may be and no doubt often has been supposed that this is an elementary matter; considering a given subject, he may answer item 1 *Yes* and item 2 *Yes*, item 1 *Yes* and item 2 *No*, item 1 *No* and item 2 *Yes*, or item 1 *No* and item 2 *No*; the four ways are equally likely, ergo positive association (*Yes-Yes*—we shall disregard, for the most part, the negative association *No-No*, and confine the term *positive* to items answered in the maladjusted manner) will occur in one-fourth the subjects by chance, and an association will be significant in proportion as it deviates from 25%. The difficulty, however, is that four ways are not equally likely. The usual understanding to that effect appears to be due to a misconception concerning the logic of probability; in what follows I shall rely upon the classic treatise of Keynes (6). Probability owes its magnitude solely to our relevant knowledge, and an alteration of the latter alters the probability accordingly. When we speak of the probability that a coin will fall heads, we implicitly include in our relevant knowledge an item of primary importance which we have not obtained from this particular coin, but from assumption, or at most from a more or less extensive experience with other coins; this item, in its form as an assumption, is that the coin is unbiased (i.e., as likely to fall heads as tails), and in its form as a generalization from experience it is that other similar coins have been found to fall approximately as often heads as tails; and this assumption is a reasonable one in the absence of specific information about this coin. If now we throw this coin a large number of times and observe that it falls heads approximately twice as often as tails, the situation is entirely changed, in that we have an additional item of relevant in-

TABLE 10

#	80	31	49	73	94	3	187	16	90	181	150	21	92	138	152	24	151	70	145	206
80	1	3	-3	-8	-2	-2	-13	-6	-7	3	-1	-6	-3	-3	-4	-9	-10	-7	-6	
31	58	2	-7	-4	0	0	-6	-6	-5	-5	-8	5	-6	-5	3	-6	-5	-6	0	
49	55	51	-1	-1	-3	-4	-3	-2	-6	0	3	-2	4	0	-4	-1	-2	0	0	
73	60	59	52	-8	-4	-6	-10	-3	-1	-4	-6	-8	-4	-5	-2	-6	-5	-9	-7	
94	60	52	48	54	-5	-11	-10	-12	-9	-2	-5	-1	-6	-12	-5	-6	-5	-12	-14	
3	50	44	46	47	44	-21	-4	-6	-7	1	2	-3	3	-7	-1	-1	-3	-15	0	
187	50	44	47	49	50	57	-7	-4	-7	-5	1	-2	-4	-6	-2	3	-1	-19	-1	
16	60	49	45	52	48	39	42	-8	-7	-4	-12	-6	-6	-8	-10	-2	-7	-9	-10	
90	53	49	44	45	46	41	39	43	-6	0	5	-5	-5	-2	-3	0	-4	-5	-1	
181	54	48	48	43	47	42	42	41		-4	-2	2	-5	-8	-5	-3	-7	-8	-1	
150	42	46	41	44	39	33	39	37	33	37	-2	-3	-2	-2	-1	1	-2	-3	-1	
21	45	49	38	46	41	31	32	45	28	35	33	1	-6	-13	-5	-10	-4	0	-8	
92	50	36	43	48	37	36	35	39	38	31	28	30	-2	0	-5	-1	-4	-2	1	
138	47	47	37	44	42	30	37	39	38	38	33	37	33	-4	-4	-2	-5	-4	-4	
152	47	46	41	45	48	40	39	41	35	41	33	44	31	35	-5	-7	-5	-4	-9	
24	46	37	43	40	40	33	34	42	35	37	32	35	35	34	35	0	6	-2	-7	
151	51	46	40	44	41	33	29	34	32	35	30	40	31	32	37	32	-6	1	-10	
70	52	43	39	42	39	28	32	38	35	38	33	33	34	34	32	22	34	1	-3	
145	47	43	36	45	44	45	49	41	35	38	31	28	30	32	32	29	26	25	-5	
206	46	36	35	42	46	30	31	39	30	30	29	35	26	31	36	34	37	28	29	
89	34	37	31	38	33	23	22	28	25	24	30	33	25	26	29	24	27	26	20	26
148	46	42	33	31	37	31	29	35	34	36	27	34	27	32	39	26	33	32	28	28
177	42	39	35	42	39	32	32	35	33	38	31	40	27	32	37	35	32	28	28	29
195	46	39	34	35	40	29	31	40	33	35	32	39	29	35	34	30	34	27	28	31
52	41	36	34	36	45	27	32	37	29	30	29	36	27	32	32	31	36	26	28	43
78	41	40	34	43	36	34	36	37	37	33	30	33	31	29	30	29	28	29	29	31
35	41	37	29	36	31	30	31	28	26	28	29	28	28	26	27	25	34	30	25	24
124	43	36	37	37	37	33	34	34	34	35	24	27	36	25	32	27	31	27	36	27
136	35	39	36	38	32	31	32	34	30	32	31	30	31	30	33	29	34	28	27	29
202	45	37	34	37	31	30	30	38	29	36	27	34	26	30	34	37	32	25	30	32
28	42	37	36	38	34	33	33	39	34	30	29	31	27	27	32	35	27	27	33	32
33	36	36	30	36	33	28	31	33	23	28	32	33	19	26	30	23	21	27	27	26
214	39	33	31	34	37	30	33	33	29	40	25	23	25	31	32	30	24	23	29	27
91	36	31	30	34	29	28	31	31	23	29	27	27	24	23	27	26	23	23	27	23
85	42	35	33	36	36	31	33	30	31	28	28	26	29	26	28	25	31	27	28	26
161	41	29	32	37	37	33	35	34	29	31	24	23	26	24	33	26	23	27	33	23
19	34	33	32	36	30	26	30	28	21	22	26	27	29	24	26	24	23	20	19	26
57	33	33	30	35	36	35	35	32	30	35	32	27	21	28	26	24	23	25	31	24
121	29	29	26	27	24	20	18	24	24	21	28	24	21	23	22	21	24	23	16	24
180	32	28	28	33	30	31	32	27	26	31	26	27	21	24	26	21	22	20	23	19
Incidence (%):																				
	67	61	60	59	54	50	50	49	49	49	47	46	46	46	46	45	45	43	42	41

formation; and the probability relative to this increased amount of relevant knowledge is no longer .5, but .67.

In like fashion, it is not an even chance that a given subject will respond *Yes* to item 1 of the Thurstone scale, *As a child did you like to play alone?* Since 34 of the 144 Pittsburgh freshman women answered *Yes*, the probability is about .24. Item 2, *Do you usually control your temper?* was answered *No* by 33 of the same group, wherefore the probability is about .23 that another subject

TABLE 10

89	148	177	195	52	78	35	124	136	202	28	33	214	91	85	161	19	57	121	180	#
5	-7	-3	-7	-3	-3	-4	-6	2	-8	-6	-1	-4	-2	-9	-8	-2	-1	3	0	80
-2	-7	-4	-4	-2	-6	-4	-3	-6	-4	-4	-4	-1	0	-5	1	-4	-4	0	1	31
4	2	0	1	0	0	4	-4	-3	-1	-4	1	0	0	-4	-3	-3	-1	3	1	49
-4	3	-8	-1	-3	-10	-4	-5	-6	-5	-7	-5	-3	-4	-7	-8	-12	-7	1	-5	73
-2	-6	-8	-9	-15	-6	-1	-7	-2	-1	-5	-5	-9	-2	-10	-11	-4	-10	2	-4	94
6	-2	-3	0	1	-6	-3	-6	-4	-3	-6	-2	-4	-3	-6	-8	-2	-11	4	-7	3
7	0	-3	-2	-4	-8	-4	-7	-5	-3	-6	-5	-7	-6	-8	-10	-6	-11	6	-8	187
0	-7	-7	-12	-9	-9	-1	-7	-7	-11	-13	-8	-8	-6	-6	-10	-5	-9	-1	-4	16
3	-6	-5	-5	-1	-9	1	-7	-3	-2	-8	2	-4	2	-7	-5	2	-7	-1	-3	90
4	-8	-10	-7	-2	-5	-1	-8	-5	-9	-4	-3	-15	-4	-4	-7	1	-12	2	-8	181
-3	0	-4	-5	-3	-4	-3	2	-5	-1	-4	-8	-1	-3	-4	-1	-4	-10	-6	-4	150
-7	-8	-14	-13	-10	-7	-3	-2	-5	-9	-6	-9	1	-4	-3	0	-5	-5	-2	-5	21
1	-1	-1	-3	-1	-5	-3	-5	-6	-1	-2	5	-1	-1	-6	-3	-7	1	1	1	92
0	-6	-6	-9	-6	-3	-1	0	-5	-5	-2	-2	-7	0	-3	-1	-2	-6	-1	-2	138
-3	-13	-11	-8	-6	-4	-2	-7	-8	-9	-7	-6	-8	-4	-5	-10	-4	-4	0	-4	152
2	0	-9	-4	-6	-4	0	-2	-4	-12	-11	0	-7	-3	-3	-4	-3	-3	0	0	24
-1	-7	-6	-8	-11	-3	-9	-6	-9	-7	-3	2	-1	0	-9	-1	-2	-2	-3	-1	151
-1	-7	-3	-2	-2	-5	-6	-3	-4	-1	-4	-5	-1	-1	-6	-6	0	-5	-3	0	70
4	-4	-4	-4	-4	-5	-2	-13	-4	-7	-11	-5	-7	-6	-8	-12	1	-11	4	-3	145
-2	-4	-5	-7	-20	-8	-2	-5	-7	-10	-10	-5	-6	-2	-6	-3	-7	-5	-5	0	206
22	1	-4	-3	-4	2	-1	4	0	1	-1	-8	-2	0	2	4	-1	-4	-3	-3	89
27	28	-5	-8	-5	-2	-5	-2	-7	-9	-1	-1	-6	-2	-2	-3	-1	-1	3	-1	148
26	31	-6	-11	-6	-3	-6	-7	-11	-11	-9	-13	-6	-10	-4	-5	-7	2	-7	177	
26	27	29	-7	-4	-5	-6	-8	-11	-7	-7	-8	-2	-5	-4	1	-3	-5	-3	1	195
26	27	33	29	-6	-6	-5	-6	-8	-4	-7	-10	0	-12	-4	-7	-6	-1	1	52	
20	24	28	26	28	-4	-5	-11	-6	-2	-6	2	-7	-6	1	-4	-9	1	-3	78	
23	27	25	27	27	25	0	-4	-5	-1	-4	1	-6	-2	2	-4	-4	1	0	35	
18	24	28	28	26	26	22	-8	-1	-8	1	-3	-1	-5	-9	2	-5	0	1	124	
22	29	29	30	27	32	26	30	-9	-5	-4	-1	-4	-6	-3	-5	-5	2	4	136	
21	31	33	33	29	27	27	23	31	-14	-8	-15	-5	-7	-3	-1	-7	-2	1	202	
22	22	32	28	25	23	21	28	25	34	-7	-12	-2	-1	-8	1	-6	-4	-5	28	
29	22	30	29	27	26	24	19	24	28	26	-7	-1	-5	1	-1	-13	-3	-6	33	
23	27	34	23	30	18	19	23	21	35	31	23	-2	-8	-10	-2	-6	1	-5	214	
20	22	26	25	20	27	25	20	23	24	21	25	20	-4	-4	-2	-5	0	-6	91	
18	22	30	24	31	25	21	24	25	26	19	23	26	21	-7	-10	-5	1	-1	85	
16	23	24	19	23	18	17	28	22	22	26	16	27	21	24	-3	-8	0	-8	161	
20	20	24	22	26	23	22	16	23	19	17	18	19	19	26	19	0	7	1	19	
23	20	26	24	25	28	22	23	23	25	24	30	23	22	21	24	16	0	-14	57	
22	16	17	22	20	18	17	18	16	20	22	20	16	17	15	16	9	22	-2	121	
22	20	26	18	18	22	18	17	14	17	23	23	22	23	17	24	15	30	18	180	
40	40	40	40	39	39	38	38	38	38	37	36	36	35	34	34	33	33	33	33	

taken at random will also answer it No. The chance association of maladjusted responses between these two items is therefore $.23 \times .24 = .054$, i.e., 8 of the group of 144 may be expected to answer both in the maladjusted fashion through chance alone. Seven did so—in all likelihood a chance association. A partial check on the reasoning involved may be secured by observing that the chance of negative association is $.77 \times .76 = .58$, so that about 84 individuals should exhibit it; 83 did so. These results may be contrasted with $.5 \times .5 = .25$, or 36 individuals who should exhibit either positive

or negative association on the more naïve theory. The following data refer in like manner exclusively to the Pittsburgh women.

Since the number of comparisons to be made among 223 items taken 2 at a time is $223 \times 222 / 1 \times 2 = 2453$, comparison was limited to 40 items, and those 40 were selected which had the highest incidences; the values of these incidences may be read at the bottom of Table 10, which presents, in the lower half, the number of individuals who should show association between the specified items if chance alone were operative; and in the top half, the divergence of the obtained figures from these.

The question naturally arises whether among the better-than-chance associations of pairs there may not be significant triads, etc. Accordingly the following diagram (Figure 2) showing all associations 10 units (subjects) greater than chance has been drawn.

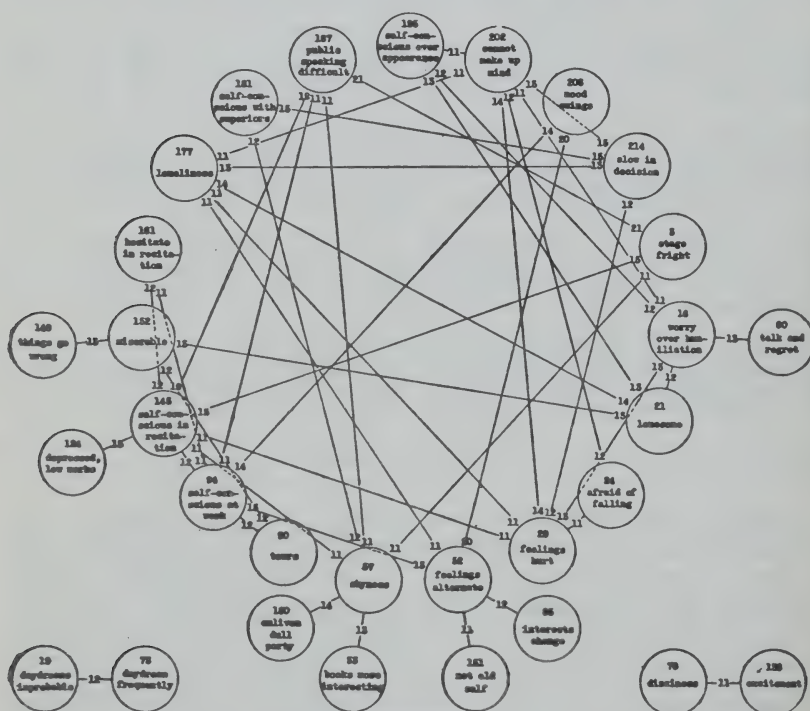


FIGURE 2

The principal inter-association among the items of highest incidence; the numbers are the excess of obtained over chance frequencies ($N = 144$) of association in maladjustment for the items indicated.

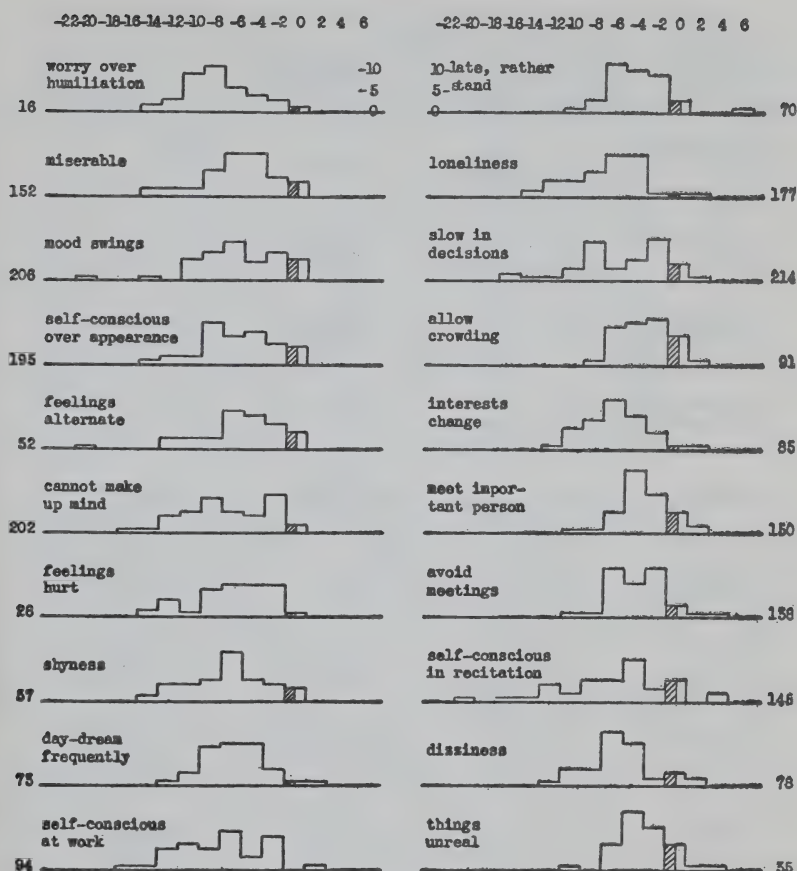


FIGURE 3

Distributions of excess of chance over obtained association of each item (of 40 items with highest incidence) with other items ($N = 144$).

The order of interrelatedness is not closely correlated with the order of incidence; the former may be measured by the sum of the bond values (representing divergence from chance in terms of number of individuals of the population of 144) or by the number of bonds involved. The former measure shows no perceptible correlation in an ordinary scattergram, and the lack of correspondence as measured by the latter may be shown by a paired list of the items in

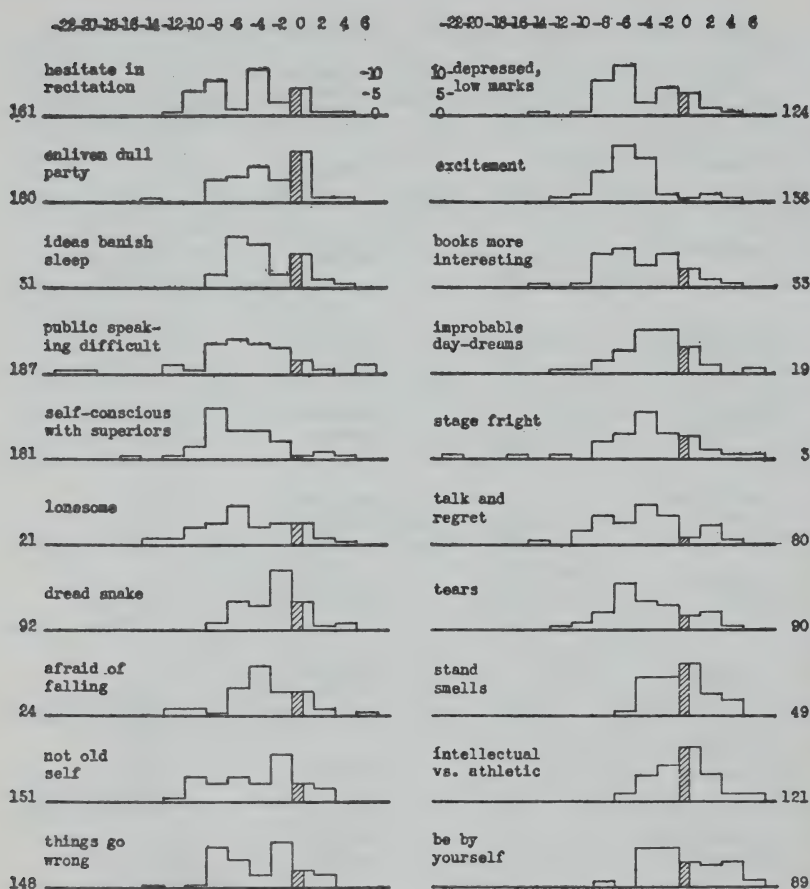


FIGURE 3 (Continued)

Figure 2 in which the incidences decrease in order as shown in Table 11.

The high degree of general cohesiveness of the scale, at least at this incidence level, may be shown by a distribution of the differences (chance minus obtained) for each item (Figure 3). The distribution of the medians by divergence classes is:

-8	-6	-4	-2	0	Total
1	17	18	2	2	40

showing clearly the tendency to cohesion.

TABLE 11

Item	Inc.	Bonds	Item	Inc.	Bonds
80	67	1	177	40	5
73	59	1	195	40	3
94	54	6	52	39	5
3	50	3	78	39	1
187	50	4	124	38	1
16	49	5	136	38	1
90	49	2	202	38	6
181	49	2	28	37	6
21	46	4	33	36	1
152	46	3	214	36	4
24	45	2	85	34	1
151	45	1	161	34	2
145	42	7	19	33	1
206	41	2	57	33	6
148	40	1	180	33	1

The material from the Beyle analysis confirms with satisfying definiteness and from a different viewpoint the findings of the Thurstone and Spearman methods. The scale, at least in these higher incidence levels, is again shown to be unexpectedly cohesive or interrelated, indicating that there is in the background a genuinely unitary psychological trait which is measured to some degree by practically all of the higher incidence items. Of special interest are the data on coherence of groups of items greater than 2 (Figure 2); these results add further strong weight to the previous descriptions of the essence of the neurotic character as characterized by inner fears and anxieties and corresponding inability to establish effective contact with the social environment. The presence of obvious duplicates (e.g., *52 feelings alternate*, and *206 mood swings*) proves to be a useful test of the method, inasmuch as these are found, as would be expected, to be associated to an extent very much greater than chance alone would explain.

A SUGGESTED REVISION

After thus examining the evidence, we are in a position to approach the problem of the selection of items for an abridgment which shall be characterized by relatively high incidence, high discriminatory power, and high cohesiveness or interassociation of items (implying natural groups) without actual duplication. We may somewhat arbitrarily assign a weight of 1 to high incidence in any

population, 2 to high discriminatory power, and 3 to satisfactory evidence of a high degree of cohesiveness; we may also allow 1 for the kind of representativeness implied in belonging to a group showing high correlation with total neurotic score. On this basis we have the following scale of weights:

Belongs to So or Fa category	1
Belongs to Darrow's neurotic constellation	3
Belongs to Thurstone's 42	2
Belongs to Harvey's 30	3 (= 1 + 2)
Incidence above 34% (highest 31) husbands	1
Incidence above 39% (highest 30) wives	1
Incidence above 37% (highest 30) Pittsburgh women	1
Belongs to associated group, Beyle analysis, with more than 1 bond over 10 (Pittsburgh women)	3
	<hr/> 15 (possible)

The following (Table 12) is a list of the 45 items having weights above 5 by the above scheme; the biserial correlation with the total score has been computed in a few representative cases, and may be seen to be substantial for these selected items.

TABLE 12

Item	<i>r</i> Bis	Item	<i>r</i> Bis	Item	<i>r</i> Bis	Item	Item
3	.30	52	.61	90		139	181
16	.65	57	.62	91		140	184
18	.22	70		94	.52	145	187
21	.54	73		98		150	195
24	.45	78		99		152	196
28	.50	79		100		161	201
30	.59	80		110		168	202
39	.73	85		124		177	206
47	.55	89		138		180	220

These are distributed thus with respect to weight totals (Table 13):

TABLE 13

Weight	No. of Items
15	2
14	2
13	2
12	5
11	3
10	8
9	1
8	2
7	5
6	15

21 lonesome may be discarded as a duplicate of *177 loneliness*; *145 self-conscious in recitation* as a partial duplicate of *161 hesitate in recitation*, and as weighting the abridgment too heavily with the peculiarities of the school situation; *187 public speaking difficult* as approximately a duplicate of *3 stage fright*; and *206 mood swings* as a duplicate of *52 feelings alternate*.

The question remains as to the most desirable number of items to retain in the revision; this will depend upon correlation with the total, reliability, printing convenience, and scoring scheme; other things being equal, the shorter scale will be preferable on the ground of the subject's convenience and cooperation. The correlations with the total for scales of the 10, 20, and 25 highest-weight items (some rather arbitrary selection being exercised when the lower limit comes within a weight class) are

10 items	.75
20 items	.84
25 items	.85

The 10-item scale is definitely too short; there is little to choose on the basis of the correlations between 20 and 25 items; both may be printed on one side of a sheet. The reliabilities (Pittsburgh women) are

20 items	.81
25 items	.83

There is also little to choose here. It is proposed, however, to substitute for Thurstone's two-response set-up, to which objection is occasionally voiced by the conscientious subject, one of five responses (0 1 2 3 4), a device which should be beneficial to the rapport and at the same time conduce to reliability and to accuracy of diagnosis. Under this scheme, the possible range of scores for 20 items will be 0-80, and the addition of five more items will increase this to 0-100. The 25-item scale has accordingly been preferred. The wording of some of the items has been slightly changed in virtue of the revised scoring procedure, and those for which a negative answer implied maladjustment have been changed in sense so that the simple sum of the item scores would yield a score proportional to degree of maladjustment.

The final revision offered is as follows:

0 means "no," "never," "not at all," etc.; 1 means "somewhat," "sometimes," "a little," etc.; 2 means "about as often as not," "an average amount," etc.; 3 means "usually," "a good deal," "rather often," etc.; 4 means "practically always," "entirely," etc.

Draw a ring around the number that describes you best.

	[Original Number]	[Weight]
Do you get stage fright? 0 1 2 3 4	3	12
Do you worry over humiliating experiences? 0 1 2 3 4	16	15
Are you afraid of falling when you are on a high place? 0 1 2 3 4	24	13
Are your feelings easily hurt? 0 1 2 3 4	28	14
Do you keep in the background on social occasions? 0 1 2 3 4	30	7
Are you happy and sad by turns without knowing why? 0 1 2 3 4	52	10
Are you shy? 0 1 2 3 4	57	12
Do you day-dream frequently? 0 1 2 3 4	73	8
Do you get discouraged easily? 0 1 2 3 4	79	9
Do you say things on the spur of the moment and then regret them? 0 1 2 3 4	80	7
Do you like to be alone? 0 1 2 3 4	89	10
Do you cry easily? 0 1 2 3 4	90	10
Does it bother you to have people watch you work even when you do it well? 0 1 2 3 4	94	14
Does criticism hurt you badly? 0 1 2 3 4	98	11
Do you cross the street to avoid meeting someone? 0 1 2 3 4	138	8
At a reception or tea do you avoid meeting the impor- tant person? 0 1 2 3 4	150	7
Do you often feel just miserable? 0 1 2 3 4	152	12
Do you hesitate to volunteer in a class recitation? 0 1 2 3 4	161	12
Are you often lonely? 0 1 2 3 4	177	13
Are you self-conscious before superiors? 0 1 2 3 4	181	15
Do you lack self-confidence? 0 1 2 3 4	184	10
Are you self-conscious about your appearance? 0 1 2 3 4	195	11
If you see an accident does something keep you from giving help? 0 1 2 3 4	196	7
Do you feel inferior? 0 1 2 3 4	201	10
Is it hard for you to make up your mind until the time for action is past? 0 1 2 3 4	202	10

Although at the date of writing only a few subjects have been secured for this revision, the indications are that it will yield a mean of about 30, a median some 5 points lower, upper and lower quartiles of about 50 and 10 respectively, and a reliability of about .98 by the

split-half method. It is intended to publish full norms later, and if possible to accompany these by checks of the validity of each item as measured by its relationship to the total score.

REFERENCES

1. BERNREUTER, R. G. The personality inventory. Stanford University, Calif.: Stanford Univ. Press, 1931.
2. BEYLE, H. C. Identification and analysis of attribute-cluster blocs. Chicago: Univ. Chicago Press, 1931. Pp. 249.
3. DARROW, C., & HEATH, L. L. Reaction tendencies relating to personality. In *Studies in the dynamics of behavior*, ed. by K. S. Lashley. Chicago: Univ. Chicago Press, 1932. Pp. 59-264.
4. HARVEY, O. L. Concerning the Thurstone "Personality Schedule." *J. Soc. Psychol.*, 1932, **3**, 240-251.
5. HOLZINGER, K. J. Statistical résumé of the Spearman two-factor theory. Chicago: Univ. Chicago Press, 1930. Pp. 43 (mimeographed).
6. KEYNES, J. M. A treatise on probability. New York: Macmillan, 1921.
7. THURSTONE, L. L., & THURSTONE, T. G. A neurotic inventory. *J. Soc. Psychol.*, 1930, **1**, 3-30.
8. THURSTONE, L. L. Multiple factor analysis. *Psychol. Rev.*, 1931, **38**, 406-427.
9. THURSTONE, T. G. The difficulty of a test and its diagnostic value. *J. Educ. Psychol.*, 1932, **23**, 335-343.

Clark University
Worcester, Massachusetts

QUELQUES TRAITS CARACTÉRISTIQUES DU QUESTIONNAIRE DE THURSTONE SUR LA PERSONNALITÉ ET UNE RÉVISION SUGGÉRÉE

(Résumé)

L'emploi du Questionnaire de Thurstone sur la Personnalité a montré une grande différence de qualité entre les divers points, et a montré aussi que l'échelle est beaucoup plus longue. On résume le travail de Harvey, de Darrow, et de Thurstone, et l'on montre que les groupes de points de toutes les régions du continuum d'incidence sauf les moins élevées ont une corrélation élevée (environ 0,83) avec l'échelle totale; les meilleurs points individuels ont des r de bi-série avec le total dans les 50 et les 60. On a divisé l'échelle en six catégories logiques, complètes, et exclusives, lesquelles ont des intercorrélations de grandeur moyenne mais moins élevées d'une façon signifiante que celles des catégories de même grandeur choisies au hasard; les analyses de facteurs de ces intercorrélations par la méthode de Spearman et celle de Thurstone s'accordent pour montrer au moins un facteur général duquel toutes les catégories sont beaucoup chargées. Les catégories ont des corrélations élevées avec le résultat total, les plus représentatives en ce sens étant celles qui comprennent des matières de fantaisie,

la mauvaise adaptation sociale, et des phénomènes hypocondriaques. La computation des inter-associations à l'excès du hasard parmi les quarante points de plus grande incidence montre une haute cohésion interne. En employant un système d'appréciation pour évaluer la conformité composée des points aux critères d'incidence, de pouvoir diagnostique, d'état représentatif, et de cohésion (sans duplication), on a choisi 25 des meilleurs points pour une échelle abrégée. On a combiné ceux-ci avec un appareil de réponse à cinq points, et les épreuves préliminaires indiquent une tendance centrale d'environ 30, les quartiles supérieurs et inférieurs d'environ 5 et 10 respectivement, et une cohérence "split-half" bien au-dessus de 0,90.

WILLOUGHBY

EINIGE EIGENSCHAFTEN DER THURSTONE'SCHEN KARAKTEREIGENSCHAFTENLISTE (PERSONALITY SCHEDULE) UND VORSCHLÄGE FÜR EINE REVISION

(Referat)

Durch Arbeit mit der Charaktereigenschaftenliste von Thurstone ist erwiesen worden, dass grosse Ungleichheit zwischen den verschiedenen Bestandteilen besteht und dass die Leiter (Skala) (scale) auch von unbequemer Länge ist. Es wird die Arbeit von Harvey, Darrow, und Thurstone zusammengefasst, und es wird erwiesen, dass Gruppen von Bestandteilen aus allen Teilen der Leiter mit Ausnahme des niedrigsten Teiles eine hohe Korrelation mit der Skala als Ganze liefern (ungefähr .83); die besten einzelnen Bestandteile lieferten doppelreihige (serial) Korrelationen in den 50 und 60ern mit der Skala als Ganze. Die Skala ist in sechs logische, erschöpfende und begrenzende Kategorien geteilt worden, die Korrelationen liefern welche von mittlerer Grösse sind aber bedeutend niedriger als die, die zwischen Kategorien von ähnlicher Grösse die aufs Geratewohl gewählt worden sind bestehen. Analyse der Bestandteile dieser Interkorrelationen mit den Methoden von Spearman und Thurstone offenbaren übereinstimmend wenigstens einen allgemeinen Bestandteil mit dem alle Kategorien stark gesättigt sind. Diese Kategorien liefern hohe Korrelationen mit der Gesamtzahl (total score). Zu den Kategorien die in dieser Beziehung die typischsten sind gehören die, die Fantasiestoff, schlechte soziale Anpassung, und hypocondrische Anzeichen in sich einschliessen. Die Berechnung der Stärke der Assoziationen der Bestandteile unter einander (inter-association) im Überschuss über die durch den Zufall verursachbaren, (in excess of chance) weisen auf eine starke innere Kohesion unter den vierzig Bestandteilen die sich am häufigsten zeigen, (of highest incidence) hin. Unter Gebrauch eines Bewichtungungsverfahrens (weighting scheme) zur Bewertung der Gesamtübereinstimmung (composite conformity) der Bestandteile mit den Kriterien der Häufigkeit, Verwendbarkeit bei der Diagnose (diagnostic power), Typischkeit, und Kohesion (ohne Duplikation) wurden 25 der besten Bestandteile für eine verkürzte Skala ausgewählt. Diese sind mit einem fünf-gradigen Beantwortungsverfahren (five-point response device) vereint worden, und vorläufige Versuche liefern eine Zentraltendenz die ungefähr 30 beträgt, obere und untere Viertelpunkte die respektiv ungefähr 50 und 10 betragen, und eine an den zwei hälften erworbene (split-half) Zuverlässigkeit die gut über .90 steht.

WILLOUGHBY

THE ISOLATION OF BLOCS IN A LEGISLATIVE BODY BY THE VOTING RECORDS OF ITS MEMBERS*

From the Department of Psychology of the University of Chicago

L. L. THURSTONE

The members of a legislative body vote on each proposition supposedly in accordance with various factors such as party affiliation, bargaining groups within the legislative body, financial inducements, lobbying, campaign pledges, and by the personal characteristics of each member. The present problem concerns merely the isolation of the principal factors that are operative in producing legislative votes, these factors to be identified by the voting behavior of the members and to be named only in so far as the factors can be associated with groups of proposals in which the factors are strongly positively loaded.

We shall assume a linear attitude continuum with an origin of neutrality or indifference and with positive as well as negative values. This continuum can be postulated for each proposal to be voted upon. A member of the voting body who favors a proposal will be allocated to a positive value on the affect continuum, while a member who opposes a proposal will be allocated to a negative value. We shall also assume that the members differ in the degree of affect which they experience favoring or opposing the different measures although each member is allowed only one vote in the actual count. Whenever the member votes in favor of a proposal we shall assume that his affect is positive for that proposal, but the value of his positive affect may be of any magnitude from near zero to the most intense favoring affect.

It will also be assumed that a member's vote on any given proposal can be regarded as a linear summation of the following form:

$$S_a = a_1x_1 + a_2x_2 + a_3x_3 + \dots a_nx_n \quad (1)$$

$$S_b = b_1x_1 + b_2x_2 + b_3x_3 + \dots b_nx_n$$

$$S_c = c_1x_1 + c_2x_2 + c_3x_3 + \dots c_nx_n$$

$$S_w = w_1x_1 + w_2x_2 + w_3x_3 + \dots w_nx_n$$

*Received in the Editorial Office, August 27, 1931.

in which the successive letters a, b, c , etc., refer to the proposals to be voted upon. The successive subscripts 1, 2, 3, etc., refer to independent and uncorrelated factors. Then

S_a = standard score of a member's affect toward proposal a .

S_b = standard score of a member's affect toward proposal b .

x_1 = standard score of a member's degree of identification with factor # 1.

x_2 = standard score of a member's degree of identification with factor # 2.

a_1 = loading of factor # 1 in proposal a .

a_2 = loading of factor # 2 in proposal a .

b_1 = loading of factor # 1 in proposal b .

b_2 = loading of factor #2 in proposal b .

These equations can be interpreted by the statement that the attitude of a member toward a proposal is a function of the extent to which the proposal is identified with the several factors and also of the extent to which the member is himself identified with each of the several factors. A factor is here any policy, slogan, advantage, a party, or other group characteristic which is found in the proposal and which is also a part of the voter's personality as a member of the legislative body. If a proposal is negatively identified with the first factor and if a member is positively identified with this factor, then the first term in equation (1) will be negative. If the second factor is not involved in the proposal, then the second term will be zero, no matter how strongly the member may feel about that factor. If the third factor is strongly represented in the proposal and if the member is also positively identified with this factor, then the third term will be positive. The summation gives the value of S_a and this is interpreted as the standard score of an individual member's attitude toward proposal a .

The next step in the analysis of voting records would be to prepare a four-fold table for every possible pair of the w proposals. There will be

$$\frac{w(w-1)}{2} \text{ pairs}$$

and for each pair we may calculate a tetrachoric correlation coefficient. These may then be entered into a large $w \times w$ table to show degree of association between all possible pairs of proposals. By the application of a multiple-factor method previously described (1)

we determine the coefficients a_1, a_2, a_3 , etc., for proposal a , and b_1, b_2, b_3 , etc., for proposal b , and so on for every proposal that was voted upon.

When these factors have been determined, the residual is determined for each proposal by the equation

$$a_1^2 + a_2^2 + a_3^2 + \dots a_n^2 + \rho^2 = 1. \quad (2)$$

If the residual ρ^2 is small, then the factors that have been isolated are sufficient to account for the voting on this proposal. If, on the other hand, ρ^2 is fairly large, such as .30 or .40, then the factors that have been isolated are insufficient to account for the voting on proposal a . This may be due to the existence of some additional general factors not yet isolated, or it may be due to the peculiarity of proposal a which may have little in common with most of the proposals by which the factors have been isolated. The same analysis can be made for each of the proposals in the list.

Each of the factors identifies a legislative grouping, a party affiliation, or bloc. After these factors have been isolated in the manner described, it is of interest to describe or name each factor or bloc. This is done by merely listing all the proposals which have strong positive loadings for the first factor. Another list is then made of those proposals which are strongly negatively loaded with the first factor. Inspection of these two lists should enable one to name the first factor. It may turn out to be one of the political parties or perhaps some other characteristic by which the two lists of proposals, positive and negative for the first factor, may be differentiated. The same procedure is followed in naming each of the other factors.

It is now of interest to describe each or some of the members in terms of the factors or blocs that have been isolated and named. The original equation can be written for each of the proposals as follows:

$$\begin{aligned} S_a &= a_1x_1 + a_2x_2 + a_3x_3 \\ S_b &= b_1x_1 + b_2x_2 + b_3x_3 \\ S_c &= c_1x_1 + c_2x_2 + c_3x_3 \end{aligned} \quad (3)$$

so as to represent the attitudes of one member toward each of the proposals in turn. These equations are written for three factors only, but the analysis can be readily extended to any number of factors. Since we now want to describe a particular member of the legislative body, it is the values of x_1, x_2, x_3 , that we want to determine.

Since S_a is a standard score it can be restated in terms of the corresponding gross score and the mean so that

$$S_a = \frac{V_a - m_a}{\sigma} \quad (4)$$

in which V_a = gross attitude score of a member toward proposal a and σ is the standard deviation of gross scores.

m_a = mean of the gross attitude scores of the whole legislative body toward the proposal a .

We may define the gross score unit so that $\sigma = 1$.

Substituting (4) in (3) we have

$$V_a - m_a = a_1 x_1 + a_2 x_2 + a_3 x_3. \quad (5)$$

Summing the gross scores of one member for all the proposals, we have

$$\sum_a^w V_k - \sum_a^w m_k = x_1 \sum k_1 + x_2 \sum k_2 + x_3 \sum k_3 \quad (6)$$

in which k represents all the proposals from a to w . The values $\sum k_1$, $\sum k_2$, $\sum k_3$ are all known since the loading of each factor in each proposal has been ascertained by a multiple-factor method.

Before we can determine the values x_1 , x_2 , x_3 , which are to describe a particular member of the legislative group, we must first evaluate

$$\sum_a^w V_k \text{ and } \sum_a^w m_k.$$

This can be done if we adopt the vote as a unit of measurement of the gross attitude continuum. But such a scale implies that some members have much stronger attitudes favoring any given proposal than some other members. We should then say, for example, that one member has an attitude of +4.5 votes for a proposal, another has an attitude of +0.3 votes for it, while a third has an attitude of, say, -1.4 votes against it. When their votes are actually cast, we shall assume that every positive attitude scores only +1 vote on the records and that every negative attitude scores -1 vote on the records.

The origin can be defined so that

$$\sum_a^w V_k = f_k - u_k \quad (7)$$

in which

$\sum_a^w V_k$ = algebraic sum of all votes cast by the particular member

f_k = number of "yes" votes that he has cast

u_k = number of "no" votes that he has cast.

Since the values of f_k and u_k are readily found from the records, it is clear that we also know the numerical value of $\sum_a^w V_k$.

In an analogous manner

$$\sum_a^w m_k = m_a + m_b + m_c + \dots m_w \quad (8)$$

$n.m_a$ = algebraic sum of all votes cast for proposal a , and similarly for every other issue.

$$n.m_a = F_a - U_a \quad (9)$$

in which

n = number of members in the legislative body.

F_a = total number of "yes" votes cast for proposal a in the entire assembly, and

U_a = total number of "no" votes cast for proposal a in the entire assembly.

Hence

$$\sum_a^w n.m_k = F_a + F_b + F_c + \dots F_w - U_a - U_b - U_c \dots U_w \quad (10)$$

$= F - U$ = algebraic sum of all votes cast by the entire assembly for all proposals

in which F = sum of all "yes" votes for all proposals in entire assembly

U = sum of all "no" votes for all proposals in entire assembly.

By means of equations (7) and (9) we should be able to reduce the number of unknowns in equation (6) to the number of factors that have been isolated. The only unknowns are then x_1, x_2, x_3 , etc. It should be possible to solve for these by the method of averages. Equation (6) would then be written as three summation equations for three groups of proposals. It would be best to arrange

all the proposals that have been voted upon in rank order according to the number of favorable votes that they have received. The list would then be divided into three groups with separate determinations of

$$\frac{w}{\sum_a V_k} \text{ and } \frac{w}{\sum_a m_k}$$

for each of the three groups if this is the number of factors that have been isolated. In the same manner one should determine Σk_1 , Σk_2 , etc., separately for each group. Then there will be as many summation equations in the form of equation (6) as there are unknowns x_1 , x_2 , etc., to be determined. The solution should then be possible though rather laborious, especially if it were desired to describe in this manner every member of the legislative body.

When the values of x_1 , x_2 , x_3 have been found, we can describe the particular member by noting which of his standard scores x_1 , x_2 , etc., are positive and which are negative. If x_1 is positive, it indicates that this member is likely to vote in favor of any proposal in which the first factor is prominently represented. He is likely to vote against a proposal which is conspicuous in this factor if his value for x_1 is negative. Similar reasoning applies to each of the other factors.

There are two checks of internal consistency for the procedures that have been outlined above. The first check is to ascertain whether the values for a_1 , a_2 , a_3 , b_1 , b_2 , b_3 , etc., are adequate to lock the observed tetrachoric correlation coefficients for all possible pairs of the proposals that have been voted upon. This is done by equations of the following form which have been described in the development of a multiple-factor method.

$$r_{ab} = a_1 b_2 + a_2 b_2 + a_3 b_3 + \text{etc.} \quad (11)$$

The correlation between all possible pairs of proposals can be calculated in this manner. These calculated coefficients can then be compared with the obtained tetrachoric coefficients, based on the actual voting records. A frequency distribution of these discrepancies should reveal a fairly good agreement between the calculated and the obtained coefficients. This criterion will be fairly well satisfied if the residuals ρ in equation (2) are, in general, small fractions of unity. If these residuals are not small fractions, then the above-described criterion of internal consistency cannot be well satisfied.

The convincingness of this criterion will depend on the ratio of the number of coefficients in the original data and the number which is the product of the number of proposals by the number of factors. This ratio should be as high as possible.

If it were desired to compute the values of x_1, x_2, x_3 , etc., for every member of the legislative group, then equation (1) could be written with known values so as to determine S_a, S_b , etc., for every member and for every issue. Whenever this value is positive, we would record a positive vote and whenever this value is negative we would record a negative vote. The agreement between such a table and the original table from which the coefficients were determined should be fairly close. In the nature of the case, this agreement might be expected to be fairly good in view of the fact that a very large number of coefficients a_1, a_2, b_1, b_2 , etc., and x_1, x_2, x_3 , etc., have been determined. In any event this total number should be considerably smaller than the number of votes cast by the entire legislative body on all of the proposals in order that this second set of determinations shall be convincing.

In most problems of this type it will probably be of interest only to determine the existence of certain principal factors and to describe them in terms of the loading coefficients $a_1, a_2, a_3, b_1, b_2, b_3$, etc. This should prove rather illuminating in some instances. It might be only the occasional situation that would call for the description of individual members of the legislative body in terms of the x -values described above.

As one check on the consistency of our reasoning we may write two summation equations with interpretation as follows.

If we sum equation (6) for all members, we have

$$\sum_{1 \text{ } a}^n \sum_{1 \text{ } a}^w V_k - \sum_{1 \text{ } a}^n \sum_{1 \text{ } a}^w m_k = 0. \quad (12)$$

since

$$\sum x_1 = \sum x_2 = \sum x_3 = 0.$$

By (7) we have

$$\sum_{1 \text{ } a}^n \sum_{1 \text{ } a}^w V_k = \sum_{1 \text{ } a}^n \sum_{1 \text{ } a}^w (f_k - u_k) = \quad (13)$$

$$\sum_{1 \text{ } a}^n f_k - \sum_{1 \text{ } a}^n u_k$$

$$\sum_{1 \text{ } a}^n \sum_{1 \text{ } a}^w V_k = (F - U). \quad (14)$$

By (10) we have

$$\sum_{k=1}^n \frac{w}{a} m_k = n \sum_{k=1}^n \frac{a}{w} m_k = (F-U). \quad (15)$$

By (14) and (15) it is seen that (12) is satisfied.

The present analysis is presented primarily as a suggestion of the general type of reasoning that might prove fruitful for the quantitative study of voting behavior. It is not presented as necessarily the only manner in which such an analysis can be made. It is the writer's hope that the outline here presented may not be seriously in error as far as its logic is concerned and that the general type of reasoning here described may prove to be applicable to the problem of analyzing voting behavior.

REFERENCE

1. THURSTONE, L. L. Multiple factor analysis. *Psychol. Rev.*, 1931, **38**, 406-427.¹

University of Chicago
Chicago, Illinois

LA SÉPARATION DES BLOCS DANS UN CORPS LÉGISLATIF SELON LES VOTES DE SES MEMBRES

(Résumé)

Cette étude est une application d'une méthode d'analyse de facteurs multiples récemment développée par l'auteur. Cette méthode de facteurs multiples est ici appliquée à l'analyse du comportement des membres d'un groupe législatif quand il s'agit de voter. On suppose que l'on puisse décrire ce comportement comme déterminé par un nombre limité de facteurs non corrélés. On suppose que chaque question à voter soit positivement ou négativement chargée de chacun des facteurs. On suppose que chacun des membres du groupe législatif soit positivement ou négativement identifié avec chacun des facteurs. Par exemple, supposons qu'un facteur soit l'amitié pour les ouvriers. Considérons un des membres qui est positivement chargé de ce facteur. Il est l'ami des ouvriers. Une proposition à voter est positivement chargée de ce facteur. Le produit de ces deux chargements, l'un pour le membre individuel et l'autre pour la proposition, est un des termes dans l'attitude totale du membre envers la proposition. La sommation de ces termes pour les facteurs principaux est une mesure de l'attitude du membre envers la proposition. Par une analyse de ce type il est possible de découvrir les facteurs principaux qui opèrent dans la dé-

¹A least square solution has been developed which supercedes the methods described in 1931. It has not yet been published.

termination des votes législatifs, de décrire chaque proposition en termes du degré auquel elle est chargée des divers facteurs, et de décrire chaque membre par une série de coefficients pour montrer son degré d'identification avec chacun des facteurs principaux.

THURSTONE

DIE ABSONDERUNG VON "BLOCS" IN EINER GRUPPE VON GESETZGEBERN AUF BASIS DER PROTOKOLLE IHRER BESCHLÜSSE

(Referat)

Der Verfasser verwendet in dieser Untersuchung eine Methode der Analyse mannigfaltiger Einwirkungen (multiple-factor analysis) die neulich von ihm entwickelt worden ist. Diese Analyse mannigfaltiger Einwirkungen wird hier auf das Benehmen der Mitglieder einer Gesetzgebergruppe bei der Wahl angewendet. Es wird angenommen, dass die Wahl (voting behavior) beschrieben werden kann, als ob sie durch eine begrenzte Zahl unverwandter (uncorrelated) Einwirkungen bestimmt wäre. Jede Frage, die durch Wahl abzumachen ist wird als mit jeder der Einwirkungen positiv oder negativ beladen (loaded) betrachtet. Jeder Mitglied der gesetzgebenden Gruppe wird als positiv oder negativ mit jeder der Einwirkungen identifiziert betrachtet. Zum Beispiel, nehmen wir die freundliche Gesinnung dem Arbeiterstand gegenüber als eine der Einwirkungen an. Betrachten wir einen Gesetzgeber, der mit dieser Einwirkung positiv beladen ist (has a positive loading with this factor). Er ist dem Arbeiterstand freundlich gesinnt. Ein Vorschlag, über den es zu wählen gilt, ist ebenfalls positiv mit dieser Einwirkung beladen. Das Produkt dieser zwei Ladungen (loadings)—eine für den einzelnen Mitglied der Gruppe und eine für den Vorschlag—ist ein Bestandteil der Gesamteinstellung des Mitglieds dem Vorschlag gegenüber. Die Summe solcher Ausdrücke für die Hauptbestandteile ist ein Massstab der Einstellung des Mitglieds dem Vorschlag gegenüber. Analyse dieser Art ermöglicht die Entdeckung der wesentlichen Einwirkungen bei der Bestimmung gesetzgeberischer Wahlen (legislative votes), die Beschreibung jeder Frage in Bezug auf ihre Beladung mit den verschiedenen Bestandteilen, und die Beschreibung jedes Gesetzgebers mit einer Koeffizientenserie die den Grad seiner Identifizierung mit jedem der wesentlichen Bestandteile darstellen soll.

THURSTONE

A PSYCHOPHYSIOLOGICAL ANALYSIS OF SEX DIFFERENCES*¹

From the Psychological Laboratories of Indiana University

HANNAH M. BOOK

I

INTRODUCTION

One of the topics in which psychology is extremely interested is that of sex differences, but a careful analysis of the field will show that our knowledge of this subject is both fragmentary and uncertain. According to certain opinions, sex differences, if they exist at all, are due to the influence of differences in experience and social training of the sexes. Others claim the differences are due to the operation of physiological factors, while many still contend that the existence of sex differences has not been established. It is our present purpose (1) to inspect in the briefest way possible the work done in this field, (2) to submit the problem to careful experiment in order to determine the quantitative facts regarding sex differences, and (3) to analyze and explain these facts in terms of acceptable science.

HISTORICAL REVIEW

The subject of sex differences has been studied both speculatively and experimentally. The first summary of the experimental results on this subject was made by Havelock Ellis in his book *Man and Woman* (11), published in 1894, but the work contains no original investigation. The first complete and systematic experimental study of psychological sex differences was made by Helen Thompson (Woolley) (39) in 1903. In this extensive investigation, *Mental Traits of Sex*, she made a series of tests for comparative purposes, dealing with motor ability, sensory and intellectual processes. Re-

*Recommended for publication by George S. Snoddy, accepted by Carl Murchison of the Editorial Board, and received in the Editorial Office, July 14, 1931.

¹This paper is constructed from a dissertation presented by Hannah M. Book to the Faculty of the Graduate School of Indiana University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology.

views of the literature on this subject were made by Woolley in 1910 (45) and in 1914 (46), one by Loutitt (28) in 1926, and brief but more recent ones by Allen (2) and Goodenough (17).

In order to summarize briefly the results of some of the experimental studies on sex differences in mental functions we shall divide them into two groups: (1) those dealing with measures of general intelligence, or a comparison of the sexes as to total scores made on composite tests, (2) those dealing with the separate parts of such tests, or measures of specific mental functions.

General Intelligence. In this group of studies we find a comparison of sexes based on total scores made on tests, such as the Binet-Simon and other group tests, on comparisons of school marks, and on statistics with regard to the number of advanced and retarded children in the schools. Such investigations have been made for the purpose of explaining certain noticeable differences in the achievement of the two sexes, such as a comparison of the number of eminent men and women, and the relative number of each sex in feeble-minded institutions. The comparison is usually made on the basis of central tendency and variability, according to the hypothesis suggested by Thorndike that the greater number of eminent individuals and also feeble-minded among men is due to their greater variability.

The most outstanding impression which one gains from a critical evaluation of the results of the numerous studies is the inconsistency of the findings, but in general the results from the most extensive studies point to the conclusion that there are no sex differences in general intelligence when this is determined from total scores.

Variability. The results of investigations here, too, are uncertain and contradictory, which we find is typical of the results of investigations all along the line of sex differences. Havelock Ellis, in his conclusion of greater male variability, was criticized by Pearson (30), who showed that there is no indication of greater male variability and that this assumption must be put aside until it has been demonstrated more scientifically.

From the results of the survey of the elementary schools of Utah directed by G. S. Snoddy (35) we find that the boys are more variable in some performances and girls in others. Furthermore, in all cases the *measure* of variability used is seen to play an important part in determining the results obtained.

Anatomical and Physiological Differences. Some of the theories

for sex differences have been of a biological nature, and refer largely to reproduction. Studies have been made on the effect of endocrine and various glandular secretions upon female behavior. Burt and Moore point out that variations of the reproductive organs mean variation in the development of glands—pituitary, adrenal, and thyroid—which tends to affect temperament. Their results show that many of the sex differences are the same in the various environments, ages, and nationalities. These are considered most important and assumed as innate.

Geddes and Thomson (14), in their book *Evolution of Sex*, also discuss intellectual and emotional sex differences from a biological point of view. They point out that a difference in constitution expresses itself in the distinctions between male and female, whether these be physical or mental, reducing them to differences in metabolism which have made women more anabolic and men more katabolic in physiological function. Burt and Moore present much the same point of view, namely, that the sexes are necessary correlates of each other, and that the differences pervade the whole organism.

Psychological or Mental Differences. In summarizing the results from her own extensive study of motor ability and comparing them with results of others, Woolley finds that all tests agree in showing that motor ability in most of its forms is better developed in men than in women. In strength, rapidity of movement, and rate of fatigue, they have a decided advantage. Ability to make delicate and minutely controlled movements is also slightly greater in men. In formation of new coordinations, such a card-sorting and crossing out "A's," women are superior. The results show that the ability to coordinate movements rapidly to new or unexpected stimuli is clearly greater in women.

In general, these results show that boys are superior to girls in rapidity of movements under conditions in which the direction of attention remains fixed; and that girls are superior in rapidity of motion in types of activity in which the direction of attention is constantly shifting—activities which involve rapid adaptation to changing stimuli.

There have been contradicting views on the general subject of the sensibility of the sexes, but in Thompson's extended investigation women were found to have finer discrimination in most things but the difference was slight. Havelock Ellis found that women were superior in rapidity of perception. He points out that the masculine

method of thought is massive and deliberate, typical of analytical thought, while the feminine method is quick to perceive and quick to act. The latter may fall into error, but are quick to retrieve the mistake, and are therefore thought to have greater tact.

The results in studies of association seem to indicate that men are faster in free associations, while women are better in practical systems of controlled association. Women are better in substitution tests or forming new associations and rapid associative thinking, while men are better in the use of associations already formed and slower to make adaptations to changing situations. Results also show that women are superior to men in memory at all ages, especially in rote memory. They form new associations more quickly and assimilate more rapidly.

Burt and Moore (8) found in a test of scope of attention, in the spot pattern test, and irregular dotting, that the boys were superior in both, though in the latter the girls were superior if the time interval was short. It was also pointed out that women notice more details, while men are more unobservant and impatient with details.

Results of tests in the more general aspects of judgment and reasoning have indicated that boys are superior in subjects which require analytical processes of thought.

Explanations Given for Sex Differences. While many of the results of these investigations are contradictory and we are left in doubt on many points, still it is evident that sex differences in specific traits do exist. In reviewing the explanations which have been given for these differences, we find that few have attempted to give an explanation based upon scientific or experimental data. Most of them have been theoretical deductions from results obtained. There are, in general, two possible explanations: one, that these differences are due entirely to the influence of social training and environment; and, the other, that they are due to a difference in physiological structure or function in the two sexes. Thus Geddes and Thomson conclude that the intellectual and emotional differences are correlated with deep-seated constitutional differences. The sexes are complementary, each higher in its own way. They insist upon a biological consideration of the differences underlying sex, which, they say, have been too much discussed by contemporary writers as though the known facts of sex did not exist at all, or almost as if these differences were matters to be settled by granting or withholding the franchise. Burt and Moore and a few others take a similar biological view.

Woolley criticizes the biological view of Geddes and Thomson and takes the view that the observed psychological differences of sex are due to differences in environment and can be explained as results due to differences in training and social influence brought to bear on the developing individual from early infancy to adult years. Jones (25) also agrees with Miss Thompson that the question of the intelligence of women is one of social necessity and ideals rather than inborn physiological character of the sexes. George (15) also takes the stand that the differences are superficial, temporary and traceable to local influences, and that they will vanish as environment is modified, as old suggestions cease to be made; that they are due to tradition.

Haggerty and Kempf (18) conclude that experimental study of adult sex differences has not as yet yielded any very significant results; there seems to be the conviction that there is a feminine type, but in just what it consists, they say we have as yet no measured experiment to prove. Thorndike thinks there are no significant sex differences unless it be one of variability; that differences among individuals are greater than those on the basis of sex. Hollingworth (20-23) also claims there is no reliable evidence of sex differences.

Goodenough (17), after reviewing the recent literature on this subject, concludes that mental differences between sexes are very small in comparison with the amount of variation existing between members of the same sex. Nevertheless, the consistency of the findings of so many independent investigations leaves no doubt that sex differences in mental traits do exist, and in measurable degree. She also points out that the pattern or profile of abilities which characterize the sexes remains relatively constant from early childhood to maturity. How these differences in pattern have originated is not apparent from the data at hand. They may be due purely to differences in social customs and ideals for the sexes, which have been impressed upon the individuals from infancy, or these very customs and ideals may themselves have been built up through the operation of minor sex differences in types of native ability.

II

PROBLEM AND EXPERIMENTAL RESULTS

The chief purpose of our experimental study was (1) to determine whether sex differences in certain mental traits actually exist to a measurable and reliable extent, and (2) to ascertain, if possible,

whether these differences can be explained in terms of variations in physiological structure or function in the two sexes.

Material. The tests used for obtaining data for this study were modifications of four of the Army Beta tests which had been used by Dr. G. S. Snoddy in a survey of the schools of Utah. These tests were selected because of the evident sex differences shown by the results of this survey in all grades of the public schools up to the sixth. It was thought that some valuable and suggestive information might be obtained by ascertaining whether these same differences would be found among adults. The tests were modified by increasing them in length and difficulty, standardizing them, and adapting them for testing adults. The advantage of these tests lies in their simplicity. They are easily given and easily scored and represent different types of simple activities which seem to have some significance for human reactions in general, such as giving an insight into the rate of adaptation to new situations.

The following tests were used: Test I (Maze) a modification of the Army Beta and Porteus Maze tests; Test II (Number Checking); Test III (Block Counting); Test IV (XO Patterns). The tests will be fully described later. Charts showing samples of the tests are shown in Chart 1.

The subjects used were 950 students in Indiana University, equal numbers being drawn from each sex, and ranging in age from 18 to 35. These students represented all departments and classes in the institution and therefore give a well-selected group.

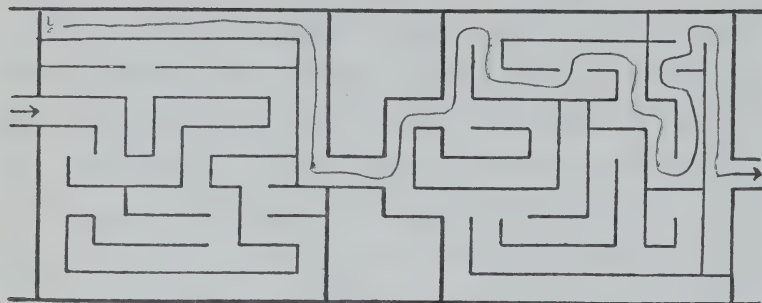
Method. The tests were all given by the writer, thus avoiding possible differences in giving, and were given under as nearly the same conditions as possible. The printed instructions for each test were distinctly read and directions carefully given.

The entire series of four tests was given to each group at one sitting with a short interval of time between the tests to avoid fatigue; the time required for the entire series was about 45 minutes.

The aim in this procedure was to make the instructions to the subject as clear as possible, and at the same time to secure the greatest possible speed and accuracy. The simplicity of the tests made it possible to secure reliable results without the influence of many factors which regularly enter into a more complex performance.

To make our method entirely clear perhaps a brief description of each test and the method of scoring the results should be given.

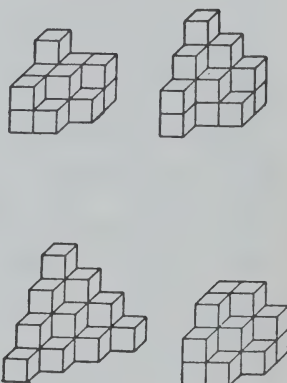
Test I



Test II

650650
041044
25792579
32813281
5519055102
3919039190
658049650849
32950173290517
6301599163019991
3900710639007106
6993108769931087

Test III



Test IV



CHART 1

CHARTS SHOWING SAMPLES OF TEST I (MAZE), TEST II (NUMBER CHECKING), TEST III (COUNTING BLOCKS), AND TEST IV (XO PATTERNS)

Description of Tests and Scoring

Test I (Maze). (Sample of test: Test I, Chart I.) In Test I there was a series of nine mazes, increasing in difficulty. A sample of this test was shown and explained and the subjects told not to turn their test blanks over until the signal "Go" was given. The instructions were then slowly and clearly read, and the signal "Get ready, go" given. The time allowed was three minutes.

Each of the first five mazes counted four points and credit was given for a half if correct. The last four mazes were given five points each according to the Porteus standardized scores for these mazes. The total possible score on this test was 40 points.

Test II (Number Checking). (Sample of test: Test II, Chart 1.) Test II contained two columns of 100 numbers increasing in their number of digits. Some of the corresponding numbers in the first and second columns were alike, others were different. A check mark was to be made on the dotted line of those that were alike and a cross when the numbers were not alike. There were 100 of these pairs to be checked. As in Test I, the directions were carefully read and the signal "Get ready, go" given.

The time allowed was five minutes. The score was the number attempted minus the number wrong. The total possible score was 100 points.

Test III (Block Counting). (Sample of Test: Test III, Chart 1.) Test III involved the counting of piles of blocks arranged in different forms. There were 45 of these piles presented in the order of their difficulty, and the problem was to estimate the entire number of blocks in as many of the piles as possible in the given time. As before, the directions were carefully read and the starting signal given.

The time allowed was nine minutes. Two points were allowed for each pile correctly estimated, making the total possible score on this test 90 points.

Test IV (XO Patterns). (Sample of Test: Test IV, Chart 1.) Test IV involved filling in blanks, containing certain patterns consisting of X's and O's arranged in patterned sequence. There were 24 of these series arranged in increasing difficulty and the problem was to complete correctly as many of these as possible in the five minutes allowed for the test. At the signal "Get ready, go" they turned their test blanks over and proceeded to fill out the blanks. Each pattern done correctly counted two points, making the total possible score 48 points.

Quantitative Results. From the frequency distributions of the 475 men and the 475 women, computations of the mean, standard deviation, standard deviation of the mean, standard deviation of the difference, and the experimental coefficient were made for each of the tests. These results are shown for each test in the following tables and graphs.

Summary and Discussion of Quantitative Results for all Four Tests. When we examine and compare the results of the two sexes in these four tests, as shown in Tables 1 and 2 and the following graphs, Figures 1 to 8, the sex differences which are shown are so large as to be beyond question. All the data have been subjected to the most rigid tests in statistical methods in order to test their reliability, and are found trustworthy.

In Test I (Maze) the results show the boys superior, the mean for the boys being 25.5 and that of the girls 20.6. The majority

TABLE 1
SHOWING FREQUENCY DISTRIBUTION OF SCORES MADE BY 475 MEN AND 475 WOMEN ON TESTS 1, 2, 3, AND 4

Class intervals	TEST 1		Class intervals	TEST 2	
	F Men	F Women		F Men	F Women
39-40	8	1	96-100	4	11
37-38	7	0	91-95	11	34
35-36	18	7	86-90	15	31
33-34	19	9	81-85	24	42
31-32	15	9	76-80	44	70
29-30	49	10	71-75	52	82
27-28	40	17	66-70	76	78
25-26	73	32	61-65	77	62
23-24	58	36	56-60	73	32
21-22	48	47	51-55	49	16
19-20	63	97	46-50	22	10
17-18	51	87	41-45	21	6
15-16	22	93	36-40	5	1
13-14	2	25	31-35	2	
11-12	1	3			
9-10	1	1	Mean	65.31	72.82
7-8	0	0	S.D.	12.40	11.95
5-6		0			
3-4		0			
1-2		0			
0		1			
Mean	25.5	20.63			
S.D.	5.82	5.20			

TABLE 1 (continued)

Class intervals	TEST 3		Class intervals	TEST 4	
	F Men	F Women		F Men	F Women
85-90	10		48-49	3	
79-84	36	6	46-47	1	4
73-78	51	12	44-45	5	4
67-72	56	27	42-43	5	12
61-66	78	58	40-41	8	13
55-60	80	63	38-39	13	23
49-54	63	85	36-37	30	31
43-48	36	71	34-35	25	41
37-42	27	71	32-33	59	74
31-36	21	46	30-31	85	69
25-30	6	17	28-29	70	79
19-24	0	6	26-27	85	69
13-18	2	3	24-25	49	28
7-12	5	5	22-23	12	13
1-6	4	3	20-21	8	6
0		2	18-19	6	4
Mean	59.48	49.40	16-17	5	3
S.D.	15.66	14.28	14-15	1	1
			12-13	1	0
			10-11	1	0
			8-9	0	0
			6-7	2	0
			4-5	0	0
			2-3	1	1
			Mean	30.07	31.37
			S.D.	5.72	5.50

of investigators stop at this point and conclude from a comparison of central tendencies that the boys are better in this ability than the girls. One may question the reliability of this central tendency and raise the question whether the difference might not be due to accidental factors such as unequal selection. To meet such criticism, and to determine whether the above difference is a real difference and not due to chance, the experimental coefficient, McCall's method for finding the reliability of a difference, was calculated and found to be 4.8 which is almost five times statistical certainty. In terms of chance an *E.C.* of 1.5 means that there is one chance out of 65,000 that the difference could be due to pure chance. Our *E.C.* for this test is more than three times this *E.C.* This is the most rigid test in statistical methods which can be applied to data. It may also be seen that the *S.D.* or measure of variability of the boys is very little greater (.62) than that of the girls.

TABLE 2

SUMMARY OF COMPARATIVE RESULTS ON SEX DIFFERENCES AS SHOWN BY EACH OF THE FOUR DIFFERENT TESTS

Number of cases—475 of each sex

Test I (Maze) Boys Superior

<i>Boys</i>		<i>Girls</i>	
Mean	25.5	Mean	20.6
<i>S.D.</i>	5.82	<i>S.D.</i>	5.20
<i>S.D.Mn.</i>	.266	<i>S.D.Mn.</i>	.238
<i>S.D. diff.</i>		.36	

Experimental coefficient 4.8, or nearly five times statistical certainty.

Test II (Number) Checking Girls Superior

<i>Boys</i>		<i>Girls</i>	
Mean	65.3	Mean	72.8
<i>S.D.</i>	11.95	<i>S.D.</i>	12.4
<i>S.D.Mn.</i>	.548	<i>S.D.Mn.</i>	.568
<i>S.D. diff.</i>		.86	

Experimental coefficient 3.14

Test III (Counting Blocks) Boys Superior

<i>Boys</i>		<i>Girls</i>	
Mean	59.48	Mean	49.4
<i>S.D.</i>	15.66	<i>S.D.</i>	14.28
<i>S.D.Mn.</i>	.718	<i>S.D.Mn.</i>	.655
<i>S.D. diff.</i>		.97	

Experimental coefficient 2.993

Test IV (XO Pattern) Girls Superior

<i>Boys</i>		<i>Girls</i>	
Mean	30.07	Mean	31.37
<i>S.D.</i>	5.50	<i>S.D.</i>	5.72
<i>S.D.Mn.</i>	.25	<i>S.D.Mn.</i>	.26
<i>S.D. diff.</i>		.36	

Experimental coefficient 1.3

These results agree with those of the Utah survey by G. S. Snoddy where the boys were found to be superior in this test for grades one to six. Porteus also found boys superior in his maze tests.

In further analysis of these results of the different test, it occurred to the writer that the difference might be referred to the assumed fact that girls are less accurate; i.e., that they attempt to answer more questions but make more mistakes.² It was therefore

²For lack of space these graphs have been omitted.

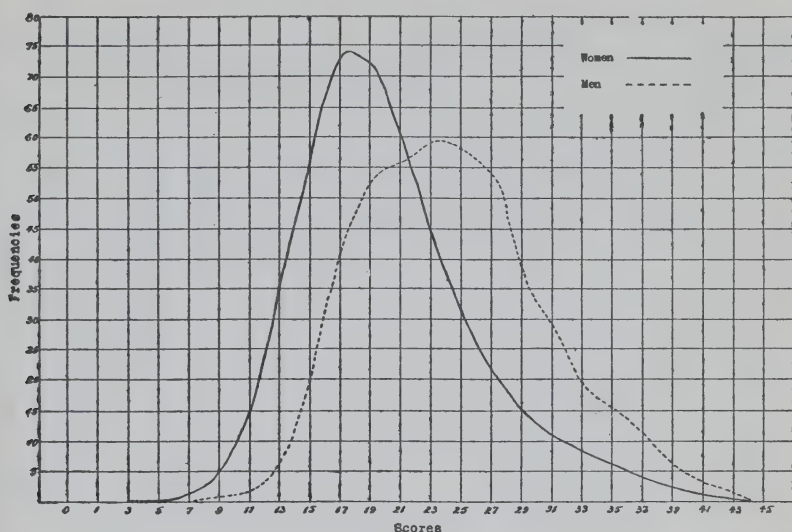


FIGURE 1

FREQUENCY POLYGONS OF SCORES OF 475 MEN (DASHED LINE) AND 475 WOMEN (SOLID LINE) IN TEST I AFTER SMOOTHING

Ordinates erected at midpoints of intervals. The scale at the left represents frequencies; the base line, scores in points.

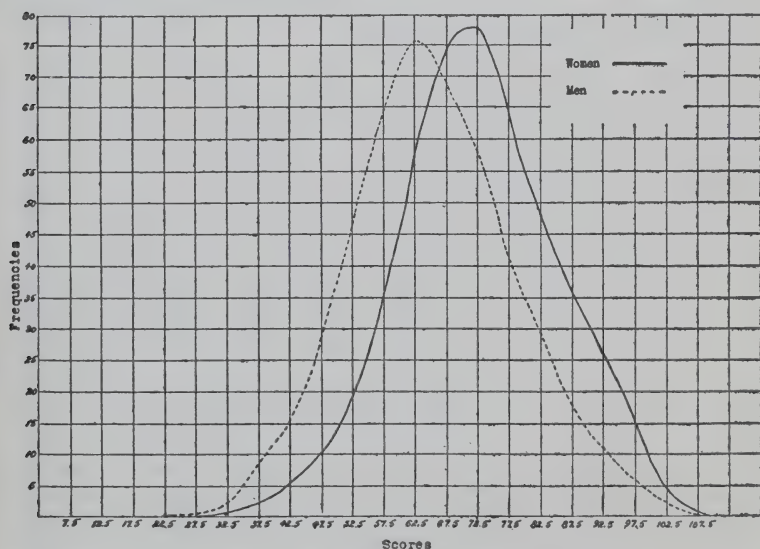


FIGURE 2

FREQUENCY POLYGONS OF SCORES OF 475 MEN (DASHED LINE) AND 475 WOMEN (SOLID LINE) IN TEST II AFTER SMOOTHING

Ordinates erected at midpoints of intervals. The scale at the left represents frequencies; the base line, scores in points.

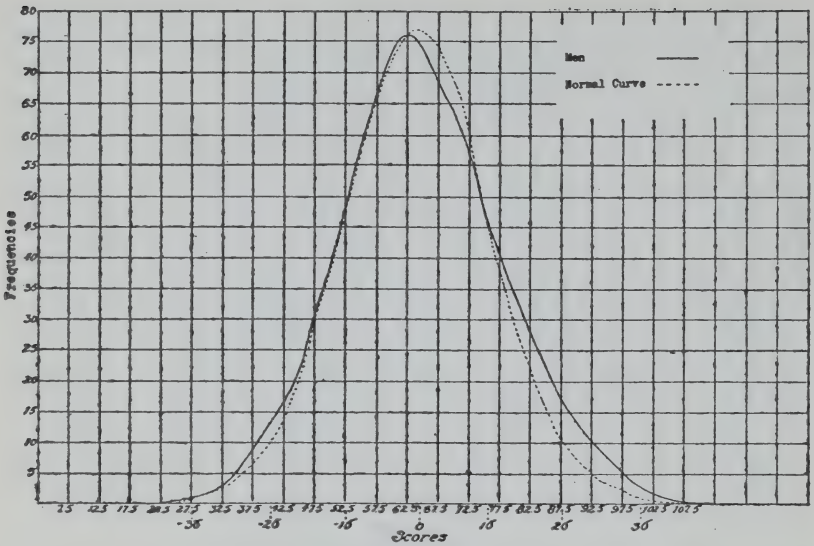


FIGURE 3

FREQUENCY POLYGON (SOLID LINE) SHOWING THE DISTRIBUTION OF SCORES OF 475 MEN IN TEST II

The dashed line shows the normal probability curve superimposed upon the actual curve. Ordinates in the actual distribution are erected at the midpoints of the class intervals. The scale at the left represents the frequencies; the base line, scores in points.

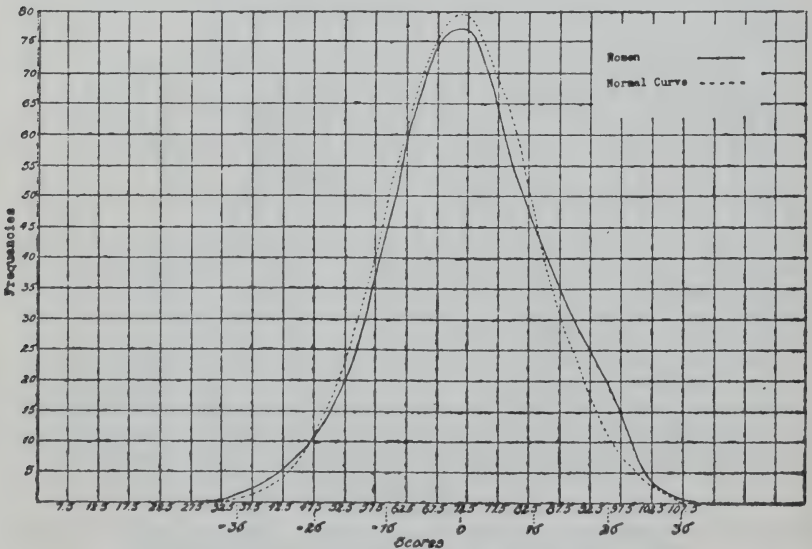


FIGURE 4

FREQUENCY POLYGON (SOLID LINE) SHOWING DISTRIBUTION OF SCORES OF 475 WOMEN IN TEST II

The dashed line shows the normal probability curve superimposed upon the actual curve. Ordinates in the actual distribution are erected at midpoints of the class intervals. The scale at the left represents the frequencies; the base line, scores in points.

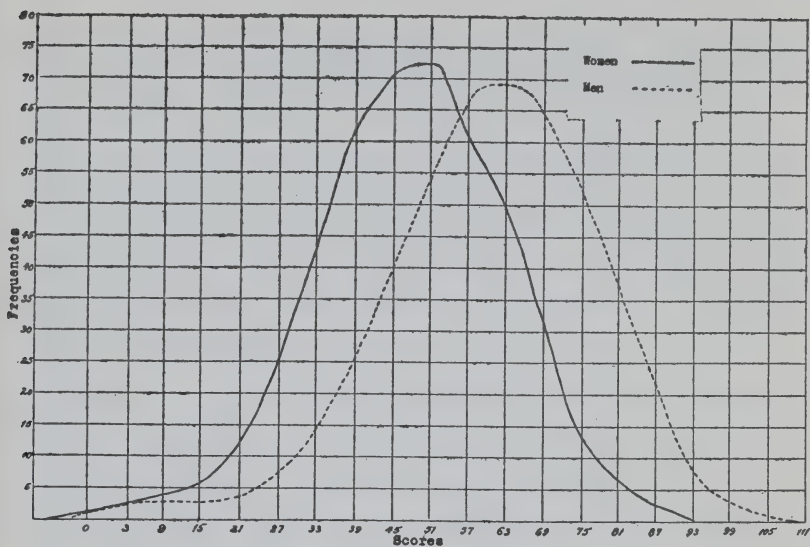


FIGURE 5

FREQUENCY POLYGONS OF SCORES OF 475 MEN (DASHED LINE) AND 475 WOMEN (SOLID LINE) IN TEST III AFTER SMOOTHING

Ordinates erected at midpoints of intervals. The scale at the left represents frequencies; the base line, scores in points.

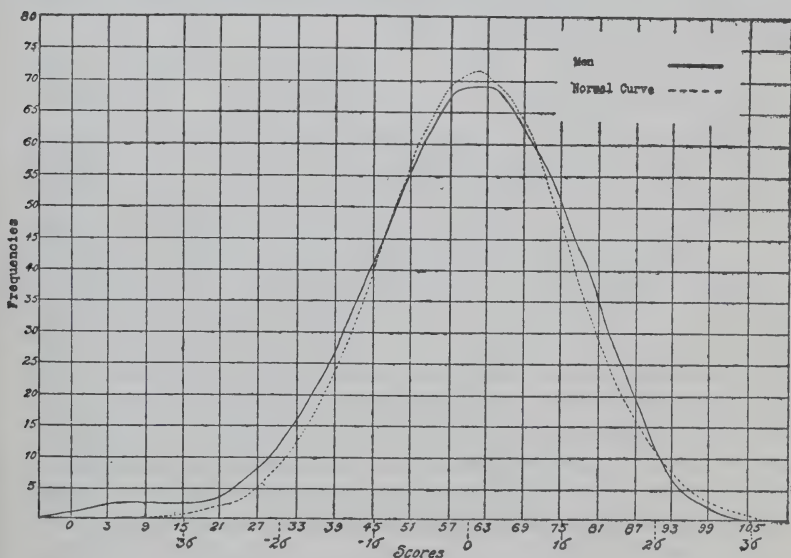


FIGURE 6

FREQUENCY POLYGON (SOLID LINE) SHOWING THE DISTRIBUTION OF SCORES OF 475 MEN IN TEST III

The dashed line shows the normal probability curve superimposed upon the actual curve. Ordinates in the actual distribution are erected at the midpoints of the class intervals. The scale at the left represents the frequencies; the base line, scores in points.

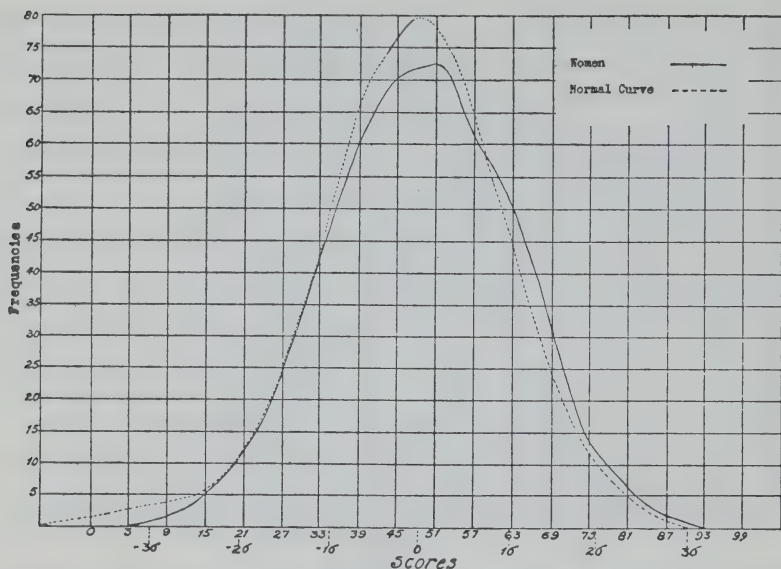


FIGURE 7

FREQUENCY POLYGON (SOLID LINE) SHOWING THE DISTRIBUTION OF SCORES OF 475 WOMEN IN TEST III

The dashed line shows the normal probability curve superimposed upon the actual curve. Ordinates in the actual distribution are erected at the mid-points of the class intervals. The scale at the left represents the frequencies; the base line, scores in points.

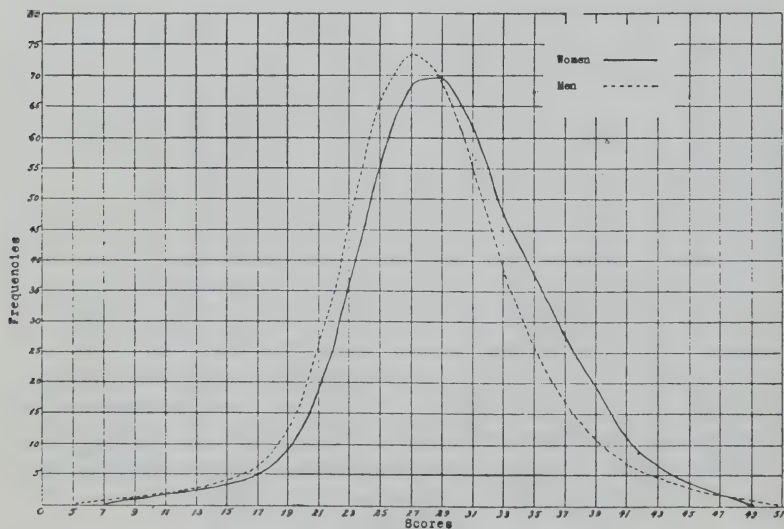


FIGURE 8

FREQUENCY POLYGON OF SCORES OF 475 MEN (DASHED LINE) AND 475 WOMEN (SOLID LINE) IN TEST IV AFTER SMOOTHING

Ordinates erected at mid-points of intervals. The scale at the left represents frequencies; the base line, scores in points.

interesting and enlightening to note that in the tests in which the men excelled they also attempted more elements than the women. And in tests in which the women excelled they attempted more elements than the men.

The results obtained in Test II (Number Checking) showed almost as remarkable a difference in favor of the women as Test I showed for the men. This again agrees with results of the Utah survey. Referring to our summary table, we see that the mean is 65.3 for the men and 72.8 for the women with a negligible difference in the *S.D.*'s. The *E.C.* for this test is 3.4 or a little more than three times statistical certainty.

Test III (Block Counting) shows the boys superior in this type of ability as was also found in the Utah survey, and the difference again is statistically reliable, as shown by the *E.C.* which is 2.998 or practically three times statistical certainty. Here again the difference in variability is negligible, being only .38.

In Test IV (*XO* Patterns), as in the Utah survey, the girls were superior in this test, but the difference is not so great here as in our other tests, or as found in the Utah results due perhaps to the fact that, as the material increased in difficulty, it assumed more the character of Tests I and III in which men are superior, as will be discussed later. Nevertheless, even though the difference is small, the *E.C.* is 1.3 which shows it is a reliable difference. This *E.C.* means that there is only one chance out of about 65,000 that our difference could be due to chance.

Conclusion. These results seem to prove definitely that sex differences do exist in measurable quantities in different kinds of activities; that the women are superior in certain types of ability and the men in others; and that these differences are not due to greater variability in either sex.

From our review of the literature on the subject we find that most investigators either claim that there are no sex differences or assume that such differences as appear are due to differences in environment and social training of the sexes. No one, as far as the writer is aware, has given any proof for this view. Some, however, think the differences may be due to physiological sex differences influencing performance, or, at least, as pointed out by Goodenough (17), that the difference in training and environment might be due to a fundamental sex difference in "drive" or interest.

It is from the latter point of view, that of fundamental innate

sex differences in physiological structure or function, that the writer will discuss the results of this study.

Our results regarding sex differences should not be interpreted as meaning that all women are superior to all men in such tests as II and IV, or that all men are superior to all women in tests of the other types. There are, of course, individual differences within any group tested; and the question is, are there demonstrable differences between the sexes?

III

PHYSIOLOGICAL BASIS AND A TENTATIVE EXPLANATION OF SEX DIFFERENCES IN MENTAL TRAITS

In the light of recent physiological studies we shall now suggest a possible explanation for these results. The theory that sex differences are due to environment and social training places the emphasis upon the differences in the kind of external stimuli to which the sexes have been subjected. If the stimuli were the only determining factors in behavior, then we could readily accept this assumption, but, as pointed out by Herrick (19), life is more than immediate reaction to stimuli experience is not something to which the organism is passively subjected. Also, as clearly stated by Child (9) individuality and individual pattern result not simply from the differential action of environment on protoplasm, but from the reactions of the protoplasm to the external differential environment. The character of this reaction must depend on the constitution of the protoplasm; the specific physiological constitution of the particular protoplasm determines the specific characteristics of the individual.

Behavior, then, in the broader view, includes every response of the living substance to stimulation. The character of these responses is determined (1) by the stimulus, a physical force acting upon excitable protoplasm, and (2) by the nature of the excited protoplasm itself. Thus it seems evident that behavior patterns of performances as indicated by our tests cannot be explained in terms of environment and social training alone.

Our results show that there are sex differences in various types of reactions. The sexes operated under exactly the same conditions when doing particular tasks, thus making the stimulus or situation at the time of the experiment the same. In order to explain these differences, we may assume that they are due either to physiological

structure or to experience. But the same differences were found by Snoddy (35) in six-year-old children as we were able to demonstrate in adults. How could experience cause the differences in these children? If they were due to experience, the differences should increase with age, but we found them to be practically the same in children as in adults. In fact, how can the experience of the sexes be such as to cause superiority of reaction to one type of stimulus in one sex and then reverse the conditions for a slight change in stimulus situation? To the writer's knowledge there has not been the slightest explanation of this. Therefore, we raise the question whether these differences could not be physiological, whether there may not be fundamentally different physiological processes, not due to experience, influencing the responses of the sexes, thus causing sex differences in different mental traits or types of performance.

Discussion and Physiological Interpretation of Results. We shall now examine the sex differences shown by the results of our four tests in the light of the findings of the leading investigations in the field of physiological psychology today. Upon analyzing the activities represented in the different tests, we see that the women are superior in those tests which require attention to details and quick adaptation to rapidly changing stimuli, as demonstrated in number checking and observation of patterns. The men are superior in the maze and block-counting tests or in activities which require a more sustained logical, and analytical attitude, where quick perceptions of detail are not as important for efficiency as slower responses and a better grasp of generalities; where a longer attention span is necessary to carry a thing through to its logical conclusion. The women represent the quick perceptual type and the men the slow-moving, exploring type. It seems probable to us that these differing capacities may be due to the physiological conditions determining the speed and character of nerve function.

It is important to note, in this connection, what has been discovered concerning protoplasmic and nervous action which might help to explain such differences. We commonly speak of protoplasm as if it were fundamentally the same substance in all organisms and all organs. But since something of the chemical and colloidal constitution of protoplasm has become known, it has become evident as pointed out by Child (10) and Ritter (34) that the body is composed not of a single protoplasm, but of a large number of specific protoplasms, each characterized by chemical processes different from any other kind.

It has long been recognized that colloids form the basis of protoplasmic structure. The fundamental constituents of protoplasm occur in a complex colloidal condition, i.e., as suspended particles in the fluid medium like drops of oil in water, bounded from the external watery medium by semi-permeable surface films. The living organism, then, consists of a mixture of substances in which a series of chemical reactions, or metabolism takes place; then we might expect to find differences in the colloidal mass or constitution of protoplasm and surface membranes of the two sexes, which is capable of influencing protoplasmic action and determining speed and character of nerve function.

Extensive experimental evidence indicates that the nerve impulse is essentially an electric phenomenon, and modern experimental data indicate that the protoplasmic limiting surfaces are the seat of electrical polarization.

According to commonly accepted views, these limiting surfaces or plasma membranes of cells in the unexcited or resting state seem to be permeable to certain or all positive ions, and impermeable or less permeable to the negative ions. Excitation or stimulation involves an increase in permeability, and consequently a more or less complete depolarization of the membrane takes place.

Changes in permeability and in electric polarization of membranes, are, according to Lillie (27), essential factors in excitation and transmission. What the stimulus does, according to Lucas (29), is to upset a condition of unstable equilibrium in the nerve membrane, i.e., a process of depolarization. Hence any difference in physiological structure which might tend to increase the permeability of the cell membranes would affect rate of functioning and thus the response.

Lillie (7) has definitely shown by means of a number of experiments on excitation and transmission that certain substances in varying amounts have a direct effect upon the permeability of the cell membranes and therefore influence excitation and transmission. He points out that the conductivity of the cell as a whole appears to vary directly with the permeability of the plasma membrane. All conditions that increase general permeability (action of cytolytic substances or unbalanced salt solutions, poisons, or other lethal agents) also increase electrical conductivity. Apparently all substances that alter the physical or chemical state of the plasma membrane, such as salts, acids, alkalies, or narcotic agents, modify the stimulating process.

The refractory period is another important factor in determining the rapidity of a response and can be explained from the point of view of membrane changes during stimulation. When a stimulus is repeated, if the second follows the first too rapidly, there will be no response. This is due to the fact that the preparation has not had time to recover or "build up" again after the depolarization of the first. During this refractory period, the tissue loses temporarily both irritability and conductivity. Lucas (29) found that the slow reactions characteristic of certain tissues are dependent on their long refractory periods, and that the length of the refractory period varies depending not upon the stimulus but upon the chemical conditions in the surroundings and the physiological state of the tissue. Here, too, was shown the influence of salts in the medium; an increase of potassium and various poisons lengthening, and an increase of calcium shortening, the refractory period. The effect of endocrine secretion on general metabolic efficiency has also been discussed by Herrick (19). Hence the refractory period is one of the important factors in determining frequency of successive impulses.

According to Lucas, impulse frequency is the fundamental concept for the theory of nerve action, the impulse frequency depending upon the relative permeability of the cell membrane and the length of the refractory period. Thus any factors influencing this permeability and the rate of recovery would affect the impulse frequency or the rate at which the individual can respond to stimuli.

Applying this to male and female activity as shown by our tests, we see that it might be possible that certain physical elements, probably certain glandular secretions or other substances acting as electrolytes, affecting the condition of the nerve centers and fibers, may tend to increase the permeability of cell membranes and produce a less rapid building-up with longer refractory periods in men. Therefore they react more slowly, allowing more time for reaction to the situation, and are able to hold things together longer before acting, which is typical of the sort of sustained activity in which our men excel. In women there is perhaps a more normal semi-permeability of cell membrane, with conditions producing a more rapid building-up and shorter refractory periods, which would account in one way for their adaptive responses to rapidly changing stimuli.

These two types of activity are demonstrated in our tests; the slow-moving sustained type of activity, typical of the kind necessary in such performances as solving the maze (Test I) and the block

counting (Test III) in which the men were found superior, and the activity required in situations demanding many quick, discreet responses and adaptations to rapidly changing conditions, as in Tests II and IV, where women excel. In our maze in which men excel, women would have a tendency to react to all points of stimulation quickly, as separate details, rather than respond to the maze as a continuous whole. They do not do so well in activities where slow responses are advantageous, i.e., where time is given to "size up" the situation. Longer refractory periods lead to perseveration with slower responses and therefore play an important rôle in systematization and generalization. In number checking, on the other hand, in the case of men, the same stimulus is repeated too quickly for the response to take place on account of their longer refractory periods and slower adaptations; while women, due to shorter refractory period and greater number of nerve impulses per unit of time, would have the ability to respond quickly to more details and be able to adapt more quickly to the changing digits.

The sex difference was not so great in our *pattern* test where the women excel. This involves initially many quick, distinct responses, but as the patterns grew longer, it gave more time to make the adaptation, since it is then necessary to hold the entire pattern in mind as a unit, and therefore this test took on something of the character of Tests I and III (Maze and Block Counting) which require more sustained attention, thus reducing the advantage of the women. In the block-counting tests again we find the tendency of the women would be to respond to each individual detail or block, quickly, without time to systematize and get a general conception of the figure. In general, if the task is fractional, i.e., divided up into parts, and the time interval short, the women are better; if, however, the task is continuous, the men are superior. It is well to keep in mind that the older conception of Lucas is inadequate to explain response to a continuous task or stimulation, since it is known that nerve substance quickly becomes adapted to a continuous stimulus and fails to respond after the first impulse aroused by the stimulus. When the nerve fiber becomes adapted, further response is impossible unless there is a change or break in the stimulus. That is why a short, quick, adequate stimulus is more effective than a slower one of equal intensity. If a stimulus of constant intensity be applied gradually there may not be any response since adaptation is taking place as rapidly as the stimulus is being applied.

For an analysis of adaptation and continuous response to a consistent stimulus, let us turn to a recent work of Adrian (1), and, on the basis of the results found by him in his recent experiments, suggest an explanation for them as applied to the two types of activity demonstrated by our tests; i.e., (1) the slow-moving sustained type, requiring slow adaptation and response to a continuous stimulus, and (2) the quick perceptual type, which requires quick adaptation and response to rapidly changing stimuli.

Adrian points out that an effective stimulus implies a certain degree and certain rate of change in the environment. The stimulus acts merely as a trigger for setting off the impulse and does not contribute the energy needed for its transmission. The impulse is a momentary disturbance which passes down the nerve fiber and is accompanied by a change of electrical potential. When an impulse reaches any point on the fiber, that point becomes refractory to external stimulation and will not transmit a second impulse until a certain time has elapsed. The result is that the stimulation which a nerve fiber can transmit must consist of one or more discrete impulses, and a continuous transmission of the excited state is impossible. In fact, the only way in which the stimulation can be made to vary at all is by variation in the total number of impulses and in the frequency with which they occur.

From this recent and important investigation by Adrian, we find again that the nature of the impulse does not depend upon the character or strength of the stimulus, but upon the local condition of the fiber; and that the number of impulses and frequencies depends upon the length of the refractory period which in turn is determined by differences in internal conditions and constitution.

He found in his experiments with nerve fibers that brief successive stimuli will bring about a succession of impulses; and if a steady stimulus of the same intensity is applied to nerve fibers, there will be only one impulse, unless the current is very strong. Adrian explains this as due to the very rapid adaptation of the nerve to the stimulus. The same process—gradual increase in adaptation—which makes a slowly increasing current unable to excite will also cause a rapid decline in the exciting power of a current which is kept at a constant value.

Now it was shown by Adrian that if a preparation be used with an end-organ or receptor attached, and a constant stimulus be applied, adaptation is much slower and a succession of impulses is produced,

while in the naked nerve fiber the excitatory process declines almost instantaneously and only one impulse is set up, unless the stimulus was very strong. Then upon comparing the behavior of the end-organ with that of the nerve fiber we find a process of adaptation comparable to that in the nerve fiber, except that adaptation is very rapid in the nerve fiber and slow in the sense-organ; i.e., it is excitable longer in the sense-organ.

Adrian's results also show that the range of discharge frequency is much the same for every type of sense-organ, as is also the duration of the refractory period and recovery. But the rate of adaptation to stimulus does vary greatly from sense-organ to sense-organ.

This difference in the rate of adaptation of the end-organs or receptors corresponds with the different types of reflex action with which they are associated. Hence the receptors may be classified, as Sherrington classified motor responses, as "postural" and "phasic." In the postural end-organs we find slow adaptation with longer excitatory processes, producing a series of impulses, characteristic of persistent or sustained responses to stimulation. In the phasic type we find quick adaptation, shorter excitation, with many quick discrete responses producing rapid fluctuating movements. Persistent activity depends on a persistent inflow of sensory impulses and the receptors concerned must, therefore, be capable of giving a persistent discharge under constant stimulation, which means they must have a slow rate of adaptation. For the phasic reflexes, a persistent discharge from the end-organ would not be necessary. The phasic reflexes are responses to sudden and rapid changes in the environment and therefore rapid adaptation is necessary.

These findings of Adrian are very pertinent to our problem. We have here an explanation of response to a continuous stimulus which is dependent upon the differentiations that have appeared in receptors. It is also to be observed that there is a difference in the adaptation rate for different sense-organs, bringing about in turn different types of reaction—the slow-moving or sustained type and the rapid fluctuating movements. These two types of activity are characteristic, as our experiments show, of the responses of the two sexes and represent differences which rest in part, at least, upon the variations in function of receptors.

In making the application of this demonstrated difference in sense-organs to sex differences in activities, we find superiority in those functions involving rapid discrete responses in girls; therefore, they

belong to the "phasic" receptor type. In those activities in which the girls excel a persistent discharge from the end-organs is not necessary, for in the phasic reflexes, or quick perceptual type, we have responses to sudden changes in the environment. This would account for that greater efficiency in all sorts of activity where stimuli are constantly changing and rapid perceptions are valuable.

On the other hand, men would be of the "postural" type, where we find slow adaptation in the receptors and therefore greater integration possible, because of the slow working of the receptors. Because of the slow adaptation, they are able to respond longer to different items, hold them more in a pattern or synthesis, thereby getting a general view of the whole situation because of this perseveration of the impulses from the sense-organs. The summated effect of a longer discharge would allow it to break through into parts of the central nervous system not accessible to brief discharges. Therefore, a larger irradiation pattern would be built up.

It is also pointed out by Adrian (1) that the only difference between sensation and impulse frequency is that the mental response rises and falls smoothly, while the nervous message consists of a series of discrete impulses; there must be a smoothing process taking place somewhere, changing the disconnected impulses into a much slower period. This again may point to the reason for the greater integrative ability in boys due to this slowing down and smoothing out of the trains of impulses into integrated wholes, as compared with the many quick discrete responses to succeeding stimuli in girls.

It was pointed out in the findings of some of the investigations that the sex differences were more pronounced in simple activities and that, as the activity becomes more complex, they tended to disappear. This might be explained by the fact that the more complex activities require a more persistent effort, and come more under the postural type; thus the girls would lose the advantage which the phasic type of receptors gives them; this does not mean, however, that they are not capable of doing this sort of thing, but simply that they lose their advantage.

IV

SUMMARY

In our experiment we have devised four test situations to study the problem of sex differences in mental functions. From the outset we saw that sex differences could be readily shown when the func-

tions employed were specific enough. In general or composite functions the sex differences were obliterated. In our four tests upon college men and women, the males were decidedly superior in two tests and the females in two; yet when the total scores for the four tests are found for each sex the sex differences are entirely negligible. This revelation of sex differences in specific mental functions opens the way for an analysis of these functions in both structural and functional terms.

In analyzing the results of these tests in the light of recent investigations, we see that the character of a response does not depend upon the stimulus originating it, but upon the particular constitution and condition of the protoplasm, which differs for different species, and no doubt for different sexes. The colloidal mass or the constituents of protoplasm may vary in different sexes due to different products of metabolism, glandular secretions, etc.

Variations in permeability of cell membranes have been shown to be affected by different substances acting as electrolytes upon the membranes, bringing about depolarization. Our assumption is that such differences in the constitution of the colloidal mass are found in the different sexes which influence the length of the refractory period, one factor in explaining the differences in their rate of reaction.

The sex differences as shown by our results cannot be explained on the basis of the variation in the experience of the sexes, since the same differences as we have shown were found by Snoddy in six-year-old children. If experience were the determining factor here, the differences should increase decidedly with age.

The message which a nerve fiber is able to transmit and the resulting response depend upon the impulse frequency; and the impulse frequency does not depend upon the stimulus but upon the conditions of the nerve fiber or synapse at the time, and upon the length of the refractory period, or rate of recovery after stimulation. It was pointed out that polarization probably takes place more rapidly in women, with shorter refractory periods; therefore, we would find a greater number of impulses per unit of time, which would account for their ability to respond quickly to more details and to rapid changes in environment; and this requires quick adaptation and short refractory periods. Men have longer refractory periods and therefore fewer impulse frequencies per unit of time, which accounts for a slow, massive, deliberate sort of reaction in men. In general, if

the task is fractional, i.e., divided up into parts, the women are better; if, however, the task is continuous, the men are superior.

The types of performance in which the women seem to excel are of the phasic or fluctuating sort, and are to be explained by the phasic form of receptor in women, in which adaptation is more rapid, as is also recovery from stimulation. This would account for the greater efficiency of women in activities where stimuli are constantly changing and rapid perceptions are valuable.

As indicated by our results, men seem to excel in the postural or continuous type of performance, and these are explained by the postural type of sense-organ in men, in which adaptation takes place more slowly, thus giving rise to longer refractory periods and a slow-moving, deliberate type of activity with sustained attention. The slowness of adaptation gives more time for the spreading of the impulse and building up of larger irradiation patterns, which would explain the greater integrative ability in men.

In looking over the activities in which each sex seems to excel, we find that these are dependent upon differences in impulse frequency, rate of adaptation, and the postural or phasic type of receptor, which we have tried to show are due to physiological differences peculiar to each sex.

REFERENCES³

1. ADRIAN, E. D. The basis of sensation: the action of the sense organs. London: Christophers; New York: Norton, 1928. Pp. 122.
2. ALLEN, C. N. Studies in sex differences. *Psychol. Bull.*, 1927, **24**, 294-304.
3. BENEDICT, F. G., & EMMES, L. E. Physiology: a comparison of the basal metabolism of the normal man and woman. *Proc. Nat. Acad. Sci.*, 1915, **1**, 104-105.
4. BOOK, W. F. The intelligence of high school seniors—a report of a state-wide mental survey. New York: Macmillan, 1922. Pp. ix+375.
5. BOYNTON, P. Sex differences. *Psychol. Bull.*, 1926, **23**, 624-625.
6. BRYAN, W. L. On the development of voluntary motor ability. *Amer. J. Psychol.*, 1892-1893, **5**, 124-204.
7. BURNHAM, W. H. Sex differences in mental ability. *Educ. Rev.*, 1921, **62**, 273-284.
8. BURT, C., & MOORE, R. C. The mental differences between sexes. *J. Educ. Ped.*, 1912, **1**, 273-284, 355-388.
9. CHILD, C. M. Individuality in organisms. Chicago: Univ. Chicago Press, 1915. Pp. x+213.

³A brief bibliography condensed from the more extensive one contained in the original study.

10. ———. Physiological foundation of behavior. New York: Holt, 1924. Pp. xii+330.
11. ELLIS, H. Man and woman. London: Scott, 1894. Pp. xvi+409.
12. FRAZIER, F. W. A statistical study of sex differences in intelligence. A.M. Thesis, Stanford University, 1918.
13. GATES, A. J. Correlations and sex differences in memory and substitution. *Univ. Calif. Publ. Psychol.*, 1916, 1, 245-250.
14. GEDDES, P., & THOMSON, J. A. The evolution of sex. London: Scott; New York: Scribner's, 1897. Pp. 322.
15. GEORGE, W. L. Intelligence of women. *Atl. Mo.*, 1915, Dec.; 1916, Jan.
16. GODDARD, H. H. Two thousand normal children measured by the Binet measuring scale of intelligence. *Ped. Sem.*, 1911, 18, 232-259.
17. GOODENOUGH, F. C. The consistency of sex differences in mental traits at various ages. *Psychol. Rev.*, 1927, 34, 440-462.
18. HAGGERTY, M. E., & KEMPF, E. J. Suppression and substitution as a factor in sex differences. *Amer. J. Psychol.*, 1913, 24, 414-425.
19. HERRICK, C. J. Neurological foundations of animal behavior. New York: Holt, 1924. Pp. xii+334.
20. HOLLINGWORTH, L. S. Variability as related to sex differences in achievement. *Amer. J. Soc.*, 1914, 19, 510-530.
21. ———. Sex differences in mental traits. *Psychol. Bull.*, 1916, 13, 377-383.
22. ———. Comparison of the sexes in mental traits. *Psychol. Bull.*, 1918, 15, 427-432.
23. HOLLINGWORTH, L. S., & MONTAGUE, H. The comparative variability of the sexes at birth. *Amer. J. Soc.*, 1914, 20, 335-370.
24. JASTROW, J. A statistical study of memory and association. *Educ. Rev.*, 1891, 2, 442-452.
25. JONES, G. E. Mental differences between the sexes. *Ped. Sem.*, 1913, 20, 401-404.
26. LASHLEY, K. S. Rats, monkeys and man.
27. LILLIE, R. S. Protoplasmic action and nervous action. Chicago: Univ. Chicago Press, 1923.
28. LOUTTIT, C. M. Bibliography of sex differences in mental traits. *Tr. School Bull.*, 1925-1926, 22, 129-137.
29. LUCAS, K. The conduction of the nerve impulse. London: Longmans, Green, 1917.
30. PEARSON, K. The chances of death. Chap. 8 in *Variation in man and woman: Vol. I*. London, 1897. Pp. 256—.
31. PORTEUS, S. D. The measurement of intelligence: six hundred and forty-three children examined by the Binet and Porteus tests. *J. Educ. Psychol.*, 1918, 9, 13-31.
32. PRESSEY, L. W. Sex differences shown by 2544 school children, on a group scale of intelligence, with special reference to variability. *J. Appl. Psychol.*, 1918, 2, 323-340.
33. PRESSEY, S. L., & PRESSEY, L. W. Further data with regard to sex differences. *J. Appl. Psychol.*, 1921, 5, 78-84.
34. RITTER, W. E. Organismal concepts. Univ. Calif. Press, Vol. XXXI

35. SNODDY, G. S., & HYDE, G. E. Mental survey of Utah schools and adaptation of the Army Beta tests. Salt Lake City: Univ. Utah, 1921. Pp. 28.
36. STAFF OF THE DIVISION OF PSYCHOLOGY OF THE INSTITUTE OF EDUCATIONAL RESEARCH, TEACHERS COLLEGE, COLUMBIA UNIVERSITY. On the variability of boys and girls from thirteen to eighteen. *Ped. Sem.*, 1926, **33**, 182-184.
37. STARCH, D. Sex differences. In *Educational psychology*. New York: Macmillan, 1919. Pp. 63-72.
38. STRONG, E. K. An interesting sex difference. *Ped. Sem.*, 1915, **22**, 521-528.
39. THOMPSON, H. B. Mental traits of sex. Chicago: Univ. Chicago Press, 1903. Pp. vii+188.
40. THORNDIKE, E. L. [Ed.] Heredity, correlation, and sex differences in school abilities. *Columbia Univ. Contrib. Phil., Psychol., & Educ.*, 1903, **11**; No. 2. Pp. 60.
41. ———. Sex in education. *Bookman*, 1906, **23**, 211-214.
42. TROLAND, L. T. The physical basis of nerve functions. *Psychol. Rev.*, 1920, **27**, 323-350.
43. WELLS, F. L. Sex differences in tapping tests. *Amer. J. Psychol.*, 1909, **20**, 354-363.
44. ———. Some properties of the free association time. *Psychol. Rev.*, 1911, **18**, 1-23.
45. WOOLLEY, H. T. A review of the recent literature in the psychology of sex. *Psychol. Bull.*, 1910, **7**, 335-342.
46. ———. General review and summaries: the psychology of sex. *Psychol. Bull.*, 1914, **11**, 353-379.

Indiana University
Bloomington, Indiana

UNE ANALYSE PSYCHOPHYSIOLOGIQUE DES DIFFÉRENCES DE SEXE

(Résumé)

Dans cette étude des différences de sexe dans les traits mentaux spécifiques, on a employé quatre tests simples pour mesurer les différents types de capacité, lesquels ont montré des différences frappantes de sexe. Le traitement statistique le plus rigide a prouvé que nos différences ont été réelles et non dues au hasard. En analysant les différences montrées dans les différents tests, on a trouvé que les femmes sont supérieures dans les activités qui exigent l'attention aux détails et l'adaptation rapide aux stimuli qui changent rapidement. On a trouvé que les hommes sont supérieurs dans les tests qui exigent une plus longue attention avec une plus lente adaptation aux stimuli constants ou persistents. Les femmes représentent le type rapide et perceptif, et les hommes le type pensif, lent à agir, et analytique. Si l'on interprète ces différences à la lumière des investigations récentes dans la psychologie physiologique, on peut les expliquer sur la base des différences dans la longueur de la période

réfractaire, la vitesse de l'adaptation, et la type postural ou phasique du récepteur. Les investigateurs antérieurs maintiennent qu'il n'y a pas de différences de sexe ou supposent que les différences qui se montrent soient dues aux différences dans le milieu ou l'entraînement social des sexes. Nos résultats montrent constamment que ces différences ne peuvent pas être expliquées par l'expérience, mais doivent dépendre de certaines différences dans la structure physiologique et la fonction particulières aux sexes. Les femmes appartiennent au type phasique où l'adaptation à un stimulus est rapide, et les hommes au type postural où l'adaptation est plus lente et la réponse à un stimulus continu est possible.

BOOK

EINE PSYCHOPHYSIOLOGISCHE UNTERSUCHUNG DER GESCHLECHTSUNTERSCHIEDE

(Referat)

In dieser Untersuchung der Geschlechtsunterschiede in Bezug auf spezielle geistige Anlagen gebrauchte man vier einfache Tests zur Untersuchung verschiedener Arten von Fähigkeiten. Durch diese Prüfungen wurden ausgeprägte Geschlechtsunterschiede enthüllt. Durch möglichst genaue statistische Behandlung der Befunde wurde ermittelt, dass diese Unterschiede tatsächlich bestanden und nicht auf Zufall beruhten. Die Analyse der in den verschiedenen Tests erwiesenen Unterschiede zeigte, dass Frauen in Bezug auf Tätigkeiten welche Aufmerksamkeit auf Kleinigkeiten und rasche Anpassung auf rasch wechselnde Reize verlangen, Männern überlegen sind. Die Männer erwiesen sich als überlegen in Tests in denen eine längere Aufmerksamkeit und eine langsamere Anpassung an konstante oder dauernde Reize in Anspruch genommen wurden. Die Frauen stellen den raschen, perzeptiven, und die Männer den nachdenklichen, analytischen Typus dar. Deutet man sie im Lichte neuerer Untersuchungen, so kann man diese Unterschiede erklären auf Basis der Unterschiede in Bezug auf die Länge der Periode des Widerstands (refractory period), die Schnelligkeit der Anpassung, und die Art der Empfindung (ob durch Lage oder durch Phase bestimmt) (postural or phasic type of receptor). Frühere Forscher machen entweder die Behauptung, es gäbe keine Geschlechtsunterschiede, oder sie nehmen an, dass die Unterschiede, die erscheinen mögen, durch Unterschiede in der Umgebung oder in der sozialen Erziehung der beiden Geschlechter verursacht werden. Unsere Befunde beweisen übereinstimmend, dass sich solche Unterschiede durch Bezugnahme der Erfahrung nicht erklären lassen, sondern von gewissen Unterschieden in der Struktur und in der Funktionierung abhängen, die jedem der Geschlechter eigen sind. Frauen gehören dem Phasentypus (phasic type) an, welcher rasche Anpassung auf einen Reiz erweist; Männer gehören dem Lagetypus (postural type) an, welcher sich langsamer anpasst und eine Reaktion auf einen dauernden Reiz ermöglicht.

BOOK

SHORT ARTICLES AND NOTES

HOW ONE RACE JUDGES ANOTHER FOR PHYSICAL ATTRACTIVENESS

RICHARD MADDEN AND LETA S. HOLLINGWORTH

The experimental studies of racial differences have been accumulating for a period of fifty years, with unusual activity in the last decade. Garth (1, pp. 233-246) has summarized over one hundred of these. An examination of this summary reveals that the interest has been primarily in ascertaining differences in mental ability. Some attention has been given to the differences in aesthetic judgment and in sensory acuity.

The present experiment is an effort to determine how one race judges the members of another for physical attractiveness. The photographs of 40 Caucasian adolescents were judged for physical attractiveness by 10 Chinese and by 10 Caucasians. Each group of judges consisted of 7 men and 3 women. All were students of Teachers College. Seven of the Chinese had been in the United States one and one-half years or less; one, two years; one, two and one-half years; and another, five and one-half years.

These 20 judges were given the following directions:

"Here are 40 photographs of New York City adolescents between the ages of 12 and 25 years. You will judge them for 'good looks' (physical attractiveness). They are to be placed into fifths of 8 each. First judge the boys and then judge the girls. You will begin by selecting the poorest for the first fifth. After they have been placed in fifths you will arrange the photographs in order between the quintiles."

When questions arose as to the meaning of good looks, reference was made to the direction sheet which specified physical attractiveness.

Tables 1 and 2 show the tabulation of the judgments given by each judge. Each photograph has been given a final rank by taking the mean of the 10 individual judgments for each race. The reliability coefficient of these composite judgments has been secured by the application of the Spearman-Brown prophecy formula to the average intercorrelation of the 10 judges of each race. The average intercorrelation of the ranks given by each judge with those of every other judge of the respective racial group was .407 for the Caucasians and .373 for the Chinese. The coefficient of reliability for the judgment of the Caucasians is .873, and for the Chinese, .856. The coefficient for the Caucasian group was also computed by correlating the sums of the ranks of the first five judges with the sums of the ranks of the second five. The correlation was $.742 \pm .049$. The coefficient

TABLE 1
SHOWING HOW THE PHOTOGRAPHS OF FORTY ADOLESCENTS WERE RANKED BY
TEN CAUCASIAN JUDGES

	A	B	C	D	E	F	G	H	I	J	Median	Mean
4c	8	11	9	26	11	8	27	21	8	8	10.0	13.7
g4c	2	15	27	18	4	15	29	7	15	29	15.0	16.1
g5c	7	5	10	4	1	10	18	9	18	27	9.5	10.9
g6c	15	22	30	33	34	29	34	31	39	32	31.5	29.9
8	21	17	20	25	38	16	15	34	13	37	20.5	23.6
10	6	7	2	1	2	2	7	1	2	5	2.0	3.5
10c	18	9	6	2	6	18	10	17	16	13	11.5	11.5
14c	5	14	28	20	36	17	20	28	33	19	20.0	22.0
g15	30	29	35	3	14	28	39	33	21	14	28.5	24.6
g15c	26	35	38	38	7	32	32	3	23	31	31.5	26.5
g17	29	13	24	24	17	7	36	27	27	12	24.0	21.6
g20c	36	24	11	35	22	39	6	37	29	6	26.5	24.5
21c	1	2	4	29	8	13	12	6	7	30	7.5	11.2
g21c	4	6	1	10	5	3	3	4	1	1	3.5	3.8
g25	39	26	8	17	19	23	24	10	19	7	19.0	19.2
g26	40	32	31	8	31	24	23	8	14	39	27.5	25.0
27	34	25	23	22	30	36	35	36	36	25	32.0	30.2
28	28	27	29	6	12	26	22	13	34	28	26.5	22.5
g29	25	36	18	13	29	31	13	22	17	4	20.0	20.8
g29c	22	30	17	14	33	20	8	16	5	9	16.5	17.4
30	38	38	34	39	39	37	40	35	37	33	37.5	37.0
31	19	21	37	15	37	19	19	18	30	18	19.0	23.3
g32	37	40	33	40	26	38	16	40	22	20	35.0	31.2
g33	16	34	12	27	28	14	21	26	38	40	26.5	25.6
34	10	19	40	7	32	6	37	14	28	2	16.5	19.5
35	17	31	39	19	24	35	30	29	24	21	26.5	26.9
36	11	8	7	36	20	27	17	2	9	38	14.0	17.5
37	20	12	22	37	18	33	25	30	25	16	23.5	23.8
38	35	39	32	21	35	40	33	32	35	34	34.5	33.6
39	24	1	5	12	3	11	28	11	10	11	11.0	11.6
40	31	37	36	28	40	34	38	39	40	35	36.5	35.8
40c	12	3	13	32	9	30	14	5	12	3	12.0	13.3
44c	13	10	16	16	10	25	9	12	31	22	14.5	16.4
45c	23	16	14	34	16	12	11	23	26	23	19.5	19.8
48c	14	28	26	11	15	4	4	19	4	17	14.5	14.2
51c	27	20	25	23	27	5	2	25	6	10	21.5	17.0
54c	32	33	19	31	13	22	26	38	32	36	31.5	28.2
56c	3	4	3	30	25	1	5	20	3	15	4.5	10.9
69c	9	18	21	5	21	9	1	15	11	24	13.0	13.4
70c	33	23	15	9	23	21	31	24	20	26	23.0	22.5

TABLE 2
SHOWING HOW THE PHOTOGRAPHS OF FORTY ADOLESCENTS WERE RANKED BY
TEN CHINESE JUDGES

	A	B	C	D	E	F	G	H	I	J	Median	Mean
4c	7	8	26	13	33	2	10	11	21	12	11.5	14.3
g4c	20	5	18	17	1	15	28	12	5	19	16.0	14.0
g5c	10	4	13	29	21	6	19	19	17	31	18.0	16.9
g6c	11	36	16	33	29	22	32	32	37	34	32.0	28.2
8	5	13	1	12	5	4	9	9	4	5	5.0	6.7
10	8	2	20	11	24	1	6	1	3	4	5.0	8.0
10c	3	1	6	7	8	7	4	27	14	10	7.0	8.7
14c	9	35	24	4	26	12	33	39	34	23	25.0	23.9
g15	1	28	5	19	10	20	12	5	2	8	9.0	11.0
g15c	36	32	33	22	16	38	26	38	33	39	33.0	31.3
g17	40	23	32	31	36	39	30	28	25	37	31.5	32.1
g20c	39	9	40	25	3	26	31	24	38	40	28.5	27.5
21c	6	29	22	1	7	10	1	6	1	2	6.0	8.5
g21c	2	16	8	15	22	16	8	10	8	9	9.5	11.4
g25	19	22	37	27	4	28	36	16	23	16	22.5	22.8
g26	21	25	11	14	9	27	3	4	18	18	16.0	15.0
27	22	6	23	39	31	13	27	17	35	33	25.0	24.6
28	31	17	19	36	19	11	24	34	24	35	24.0	25.0
g29	32	21	28	20	20	23	13	13	30	20	20.5	22.0
g29c	12	11	4	5	2	9	23	7	11	22	10.0	10.6
30	27	12	14	32	25	32	40	20	40	28	27.5	27.0
31	26	19	12	26	13	17	22	37	27	13	20.5	21.2
g32	23	15	21	34	18	30	29	29	9	25	24.0	23.3
g33	30	20	30	16	38	40	38	35	16	27	30.0	29.0
34	28	18	38	30	34	25	39	33	6	17	29.0	26.8
35	16	37	9	21	35	33	18	40	32	11	26.5	25.2
36	17	40	17	6	14	18	25	30	22	3	17.5	19.2
37	4	30	31	28	27	21	2	22	19	15	21.5	19.9
38	18	27	7	38	30	35	15	26	13	26	26.0	23.5
39	14	7	3	3	11	3	5	2	7	1	4.0	5.6
40	35	10	25	37	37	31	35	36	15	24	33.0	28.5
40c	33	34	36	24	40	34	37	21	36	38	35.0	33.3
44c	37	26	27	10	28	37	34	18	26	36	27.5	27.9
45c	15	38	10	23	32	8	14	23	39	7	19.0	20.9
48c	34	3	15	18	12	24	16	15	12	30	15.5	17.9
51c	24	14	34	8	23	36	20	25	20	14	21.5	21.8
54c	25	39	29	40	15	5	21	31	28	21	26.5	25.4
56c	29	31	35	9	17	29	7	3	31	32	29.0	22.3
69c	38	24	39	35	39	14	17	14	29	29	29.0	27.8
70c	13	33	2	2	6	19	11	8	10	6	9.0	11.0

of reliability for 10 judges would be $.852 \pm .030$ when computed in this manner. The two methods agree remarkably well.

It is evident from these correlations that the Chinese use their standard for physical attractiveness as consistently within their own group as do the Caucasians. If both groups are using the same standard we should expect the coefficient of correlation between the composite judgments to approximate that of the coefficients of reliability. However, the coefficient which is actually obtained is $.470 \pm .069$. We conclude that the standard which the Chinese use so consistently is not the same as that which the Caucasians use, although it must be similar in some respects, since a positive coefficient is obtained.

The mean correlation between the ranking of one Caucasian judge with the sum of the ranking of all the Caucasian judges is .683. The judgment of a single Caucasian more nearly approximates the composite Caucasian judgment than does the composite judgment of 10 Chinese. We conclude that the Chinese standard differs from the American standard. The nature of the difference is difficult to determine.

The Chinese were asked to name some of the factors which affected their judgments. They mentioned straight hair, a full face, the smooth eyelid, the American cinema, features rather than expression, and age.

We attempted to appraise the relative influence of some of these factors. It is very probable that good looks is a gestalt and that any effort to discover a dominating element will be futile. The same may not be true for unattractiveness, however, for a trait such as strabismus may be more potent in determining unattractiveness than the remainder of any gestalt. However, we selected two elements and submitted them to judgment. Judgments were also secured for resemblance to stereotypes. These judgments were made in the following manner.

The 10 photographs were selected which showed the greatest discrepancy in ranks for the two races. These were given a new order of rank. The first was the photograph which had the highest positive difference secured by subtracting the Chinese rank from the Caucasian rank and the tenth photograph had the greatest negative difference, secured in the same manner. Thus a scale was made from greatest to least relative preference by Caucasians, or from least to greatest preference by Chinese.

Six judges who had not previously seen the photographs were asked to rank them in order for five traits; namely, (1) resemblance to the Jewish type, (2) fullness of face, (3) resemblance to your own stereotype of a cinema-star, (4) straightness of hair, and (5) resemblance to the Chinese type. The averages of the intercorrelations among the judges is given in Table 3. With one exception, these correlations are high enough to indicate that the traits are recognized as existent. It is interesting to note that a Jewish and a Chinese stereotype are as readily discernible as the two purely physical characteristics: fullness of face and straightness of hair.

TABLE 3

THE CORRELATIONS OF JUDGMENTS UPON FIVE TRAITS BY SIX JUDGES

Trait	Average of the intercorrelations of six judges	Reliability of the combined judgment of six judges
1. Resemblance to Jewish type	.293	.713
2. Fullness of face	.323	.741
3. Resemblance to stereotype of cinema-star	.065	.295
4. Straightness of hair	.277	.697
5. Resemblance to Chinese type	.329	.746

The 10 photographs were given a rank order in each of the five traits on the basis of the sums of the ranks of the judges. This order of merit was correlated with the order for Caucasian preference. Table 4 gives the ranks and the correlations.

Only the correlation between the cinema-type and the Caucasian preference is great enough to be of any significance. It seems surprisingly high when we consider the unreliability for the judgment of this trait. This relationship should not be misconstrued to mean that there is a high correlation between the Caucasians' judgment of good looks and their judgment of resemblance to the cinema-type. It means simply that an individual who is placed several steps higher on the scale of good looks by the Caucasians than by the Chinese is much more likely to resemble the Caucasian stereotype of a cinema-star than is an individual who is

TABLE 4

THE RANK OF THE COMBINED JUDGMENT OF SIX JUDGES FOR FIVE TRAITS

	Preference by Jewish Caucasian	Jewish type	Full face	Cinema- type	Straight hair	Chinese type
8	1	9	4	1.0	1.0	4.5
g15	2	8	9	2.5	9.5	7.0
g26	3	6	5	2.5	3.0	4.5
70c	4	1	1	5.5	7.5	2.5
38	5	10	8	7.0	9.5	10.0
g17	6	5	10	4.0	4.0	8.0
56c	7	4	2	8.0	6.0	1.0
44c	8	3	6	5.5	7.5	9.0
69c	9	2	3	9.0	5.0	2.5
40c	10	7	7	10.0	2.0	6.0
rho		— .430	— .030	+ .891	— .064	— .091
P.E.		± .182	± .223	± .046	± .222	± .222

ranked several steps higher on the scale of good looks by the Chinese than by the Caucasians. The correlation would be numerically the same but negative if we should substitute the Chinese preference for the Caucasian preference. This would be meaningless because the ranking for cinema-type is for Caucasians which serves in no way whatever to predict what might occur if we had a Chinese rating.

There is a tendency for the Chinese preferences to be related to the Jewish type. Racial prejudice may have been operative in the Caucasian judgments.

No attempt was made to account for the community of judgment of the two racial groups. The extent of this similarity is indicated by the correlation of .470. This correlation could be produced easily by common elements of good looks in the racial groups. Such qualities as proportionate features are quite likely to be recognized by every racial group unless a disproportionate figure secured by artificial treatment is traditional for one of the races involved in the comparison. It must not be forgotten that some of the Chinese had been in the United States for a few years. All of them had been here for at least a short period of time during which the concept of good looks was being developed no doubt through the ordinary channels of communication. Even before their departure for the United States, the Chinese were not completely isolated from Caucasian contacts. It seems that there are several factors which might account for the similarity. Our interest lies in the fact that there is some similarity in the judgments, although not enough to justify the statement made by Garth: "Beauty seems a common coin which passes as currency in any country or province" (1, p. 234). The type of beauty exhibited by photographs does not fall into this generalization. The first choice of the Caucasians is the tenth choice of the Chinese and the first choice of the Chinese is the thirty-third choice of the Caucasians. Physical attractiveness seems to possess certain elements which are common to both Chinese and Caucasians, and other elements which are characteristic of each race.

SUMMARY

The photographs of 40 Caucasian adolescents were judged for physical attractiveness by 10 Caucasians and by 10 Chinese.

The average intercorrelation of each Caucasian judgment with the other 9 Caucasian judgments is .407; and the average intercorrelation of each Chinese judgment with the other 9 Chinese judgments is .373. The reliabilities of the Caucasian and Chinese judgments are .873 and .856, respectively.

The correlation of the composites of the two judgments is .470.

Efforts to determine the relationship of five factors to the discrepancies between the judgments of the racial groups were largely fruitless, with the possible exception that the Caucasian stereotype of a cinema-star is related to Caucasian preference.

The judgments revealed that Jewish and Chinese stereotypes are as readily discernible as are the physical qualities of straightness of hair and fullness of face.

REFERENCE

1. GARTH, T. R. *Race psychology*. New York: Whittlesey House, 1931. Pp. 260.

*State Teachers College
Indiana, Pennsylvania*

*Teachers College
Columbia University
New York City*

DEVELOPMENTAL AGE AND ADOLESCENCE

J. EDWARD RAUTH AND PAUL HANLY FURFEY

It is generally assumed in the literature on adolescence that the behavior changes which characterize this period have a physical basis. In other words, the physiological phenomena of puberty are supposed to cause the psychological phenomena associated with adolescence. Yet, in spite of the evident importance of this assumption, few attempts have been made to test it out by scientific methods.

In 1926 one of us (2) published a graph showing an interesting parallelism between the ages of boys leaving a troop of Boy Scouts and the age of puberty as reported by other investigators. The implication, of course, was that puberty was responsible for the boys' loss of interest in Scouting. The following year Lehman (7) discovered independently a similar parallelism in girls between puberty and loss of interest in doll play. The same method has since been used to study the effect of puberty on general play activities (3, 8) and on vocational choices (9).

A slightly different approach was made by Leal (6). She studied a group of 1876 boys and 2294 girls, Grades 4 to 12, from the public schools of New Britain, Connecticut. These children were rated as to physiological maturity on a five-point scale by two physicians. At the same time they were rated as to 17 personality traits by their grade teachers. The results of the ratings were tabulated and from an inspection of the tables Leal endeavored to pick out the traits associated with physiological adolescence.

In spite of the suggestive nature of the above investigations, they labor from certain common defects:

1. None of these studies succeeds in eliminating the effects of chronological age. Leal's study is the only one which even attempts to do so. Even in this case the attempt is, perhaps, not entirely successful since it is accomplished by inspection of the tables rather than by statistical methods. The other investigations are based directly on chronological age and they measure physiological age only by assuming that it bears a certain relationship to chronological age.

2. None of these studies employs a satisfactory measure of maturity of behavior. They are either based on single traits (2, 3, 7, 8, 9), usually reported on a questionnaire (3, 7, 8, 9), or else on an unstandardized rating scale (6). In no case is there any attempt to determine either the validity or the reliability of the instrument used to measure behavior.

3. None of the studies employs statistical methods for determining the closeness of the relationship between behavior and maturity. In each case the relationship is merely estimated by the inspection of graphs or tables.

In the present investigation an attempt was made to avoid some of the above defects. As subjects, 84 boys from an institution for dependent children were employed. Of these 33 were 13 years old, 32 were 14, and 19 were 15. In order to eliminate as far as possible the influence of chronological age, each of these groups was studied separately.

Maturity of behavior was measured by the revised scale for developmental age (4) already published by one of us. This scale has been standardized for the precise purpose of measuring maturity of behavior ("developmental age"). It is reported to have a reliability of .94 at age 13, .93 at age 14, and .94 at age 15. It ought therefore to give a good estimate of the subjects' maturity.

Physiological maturity was estimated by Crampton's (1) norms. All the subjects were classified into two groups, the one comprising all those who were prepubescent or pubescent by these norms, the other comprising all those who were postpubescent. Although the Crampton norms are admittedly imperfect, they are still the best available way of measuring puberty in boys.

Biserial r 's were then calculated on the Kelley-Pearson formula (5, p. 248) between maturity of behavior as measured by the scale for developmental age and puberty measured as described in the last paragraph. The results are shown in Table 1.

The relationship between puberty and maturity of behavior (developmental age) as measured by score on the entire scale varies from $-.16$ at age 13 to $.79$ at age 15. The extreme difference in the size of these coefficients is probably due merely to chance; but it might conceivably be due, at least in part, to the fact that among the 15-year-old adolescents puberty had been established longer than among the 13-year-old adolescents and consequently had had more time to affect conduct.

The difference in size among the correlations between puberty and the separate tests of the scale are probably too small to be significant. Taken at their face value, however, they would indicate that the relationship with puberty is closest in the case of Test 6. ("Things to Think About"), followed in order by Test 1 ("Things to Do"), Test 3 ("Books to Read"), Test 2 ("Things to Be When You Grow Up"), and Test 5 ("Things to See"). The smallest coefficient was yielded by Test 4 ("Things to Have").

TABLE 1
CORRELATION BETWEEN TEST RESULTS AND PUBERTY AT SPECIFIED AGES
(Biserial r 's)

	Age 13	Age 14	Age 15	Mean
Test 1	.06	.32	.49	.29
Test 2	— .37	.22	.73	.19
Test 3	— .15	.11	.73	.23
Test 4	— .17	.17	.38	.13
Test 5	— .22	.23	.46	.16
Test 6	.18	.07	.80	.35
Entire scale	— .16	.42	.79	.35

It was felt that the influence of chronological age was sufficiently eliminated by calculating the coefficients separately for each age group. But to determine whether this was still a disturbing factor, partial r 's (chronological age constant) were calculated between puberty and maturity as measured by the entire scale. By this method coefficients of —.16, .43, and .80 were obtained for ages 13, 14, and 15 respectively. These differ so little from the raw coefficients of —.16, .42, and .79 that it was not considered worth while to calculate partial r 's for the separate tests.

CONCLUSION

With chronological age constant, there seems to be a slight positive relationship between physiological maturity as measured by the Crampton norms and maturity of behavior as measured by the developmental-age scale.

REFERENCES

1. CRAMPTON, C. W. Physiological age—a fundamental principle. *Amer. Phys. Educ. Rev.*, 1908, **13**, 141-154, 214-227, 268-283, 345-358.
2. FURFEY, P. H. The gang age; a study of the preadolescent boy and his recreational needs. New York: Macmillan, 1926. Pp. xii+201.
3. ———. Pubescence and play behavior. *Amer. J. Psychol.*, 1929, **41**, 109-111.
4. ———. A revised scale for measuring developmental age in boys. *Child Develop.*, 1931, **2**, 102-114.
5. KELLEY, T. L. Statistical method. New York: Macmillan, 1923. Pp. xi+390.
6. LEAL, M. A. Physiological maturity in relation to certain characteristics of boys and girls; a study of school children in New Britain, Connecticut. Philadelphia: Univ. Pa., 1929. Pp. 118.
7. LEHMAN, H. C. A study of doll play in relation to the onset of pubescence. *J. Genet. Psychol.*, 1927, **34**, 72-76.
8. LEHMAN, H. C., & WITTY, P. A. A study of play in relation to pubescence. *J. Soc. Psychol.*, 1930, **1**, 510-523.

9. ———. A study of the vocational attitudes in relation to pubescence. *Amer. J. Psychol.*, 1931, **43**, 93-101.

Catholic University of America
Washington, D. C.

A STUDY OF THE EMOTIONAL CHARACTERISTICS OF WOMEN PRISONERS

K. C. GARRISON

In this study 138 of the inmates of three penal institutions for white women in North Carolina were used as subjects. The chronological age range of these prisoners was from 15 to 30 years, and each case had committed some offense serious enough to be classified by the court as a misdemeanor. Thus, every girl in these institutions between 15 and 30 years of age who had committed a misdemeanor, save some few who were paroled during this study and several others not available due to the prison's program, was used in this study. The girls, with the exception of one Italian girl, were all native-born Americans. The chief offenses committed by these girls were: sex, 51%; delinquency, 18.1%; larceny, 5%; and robbery, 3.6%.

The *Thurstone and Thurstone Neurotic Inventory*, in the form of a questionnaire (4), was used for the measurement of emotional instability of these 138 subjects. Those questions which, from former results, showed little or no differentiation were omitted. Leaving these out would affect the reliability of the test and scores very slightly, if any.

The Thurstone and Thurstone list contained 223 questions. But, after these non-differentiating questions were omitted, the revised inventory consisted of 182 questions.

Describing the neurotic inventory, Thurstone says:

"It was compiled mostly from several shorter lists already published, including Woodworth's psychoneurotic inventory, House's monograph on this subject, Laird's schedules of questions, Freyd's list of questions about introversion and extroversion, and Allport's list of questions about ascendance-submission. Professor Woodworth's list is so well edited that we have not succeeded in making any improvement on it except for the increase in the length of the list whereby its reliability is raised" (4, p. 3).

Ladu and Garrison in 1931 wrote:

"The total score on the neurotic inventory is the total number of unfavorable or maladjusted answers that the subject returns. A numerically high score represents, therefore, an emotionally unstable personality. A numerically low score indicates the absence of emotional strains or worries, and thus indicates an emotionally well adjusted personality. The maximum possible score was, of course, 182, since that would be the score if

every one of the questions were answered in an unfavorable way.

"The scores obtained on the neurotic inventory, that is, the number of unfavorable responses per girl range from 25 to 121. A comparison of the scores made by this group with those obtained from 694 students from the freshman class at the University of Chicago will show that this group made an average emotional instability score of 65.67 which is 21.85 points higher than the average made by the college freshmen. This was true even in the light of the fact that some of the questions that were of little value in testing emotional factors were eliminated from the test for this study of delinquent girls, thus probably lowering slightly their average emotional instability scores" (2).

Aside from the number of neurotic responses so outstanding among delinquent girls, there are other questions in the neurotic inventory which, when grouped together, tend to bring out specific traits of individuals which are of interest in making a more careful qualitative analysis of these delinquents. A comparison of the unfavorable responses on the various questions was made with those made by Mathew's group of unselected girls on the same question and found in Mathew's *Adaptation of the Woodworth Scale for the Measurement of Emotional Stability* (3).

A study of these comparisons as presented in Table 1 will show a large ratio of neurotic responses on the part of the women prisoners in each group of questions. The conflicts with the home as revealed in their response to the questions *Did you ever have a strong desire to run away from home?* and *Has your family always treated you right?* show striking neurotic tendencies. This is directly in harmony with the frequency of response given these questions by 92 men prisoners at San Quentin Prison, California (1). There was a strong desire to commit suicide present among these women prisoners, another neurotic tendency similar to the one found in the study referred to with men prisoners.

The abnormal fears ran rather high for this group, which is directly contrary to the fears found among men prisoners in California. Aside from the great fear of lightning and fire, these women prisoners differed very little from the unselected girls in their general fear habits or impulses.

Another characteristic of delinquent girls revealed by this test is the number of their physical ailments. In these they differ considerably from the unselected group. They have strange feelings; Their eyes seem to pain a great deal. They feel suffocated and bored most of the time. Their sleep is disturbed and they do not feel rested in the morning. However, in spite of these disturbances, they seem to tire less easily than unselected girls, have fewer bad headaches, and not so many pains.

The delinquent girls also exhibit a pronounced tendency toward morbid depression, if the findings of this table are presumed to be an indication. They are unhappy and gloomy, and their feelings alternate between happiness and sadness without apparent reason.

TABLE 1

PERCENTAGES OF SYMPTOMATIC RESPONSES TO CERTAIN QUESTIONS IN THE THURSTONE AND THURSTONE NEUROTIC INVENTORY, ALONG WITH A COMPARISON OF THE RESULTS FOR DELINQUENT AND UNSELECTED GIRLS

Questions	Delinquent girls Percentage	Unselected girls Percentage	Ratio of neurotic responses of delin- quents to those of non-delinquents (unselected girls)
<i>Unsociable traits</i>			
Like to play alone as child?	28.0	19.0	1.474
Like to be by yourself?	51.0	15.0	3.400
Happy when fourteen to eighteen?	34.0	14.0	2.429
Get tired of people quickly?	20.0	11.0	1.818
<i>Conflicts with home and persons</i>			
Strong desire to run away from home?	60.0	4.0	15.000
People find fault with you more than you deserve?	50.0	13.0	3.846
Do people think you are selfish?	30.4	16.0	1.900
Often feel lonesome, even when with others?	64.0	30.0	2.133
Feel not satisfactorily adjusted to life?	47.0	30.0	1.567
Family always treated you right?	48.0	1.0	48.000
Think you are regarded as queer?	35.0	30.0	1.167
<i>Abnormal fears</i>			
Uneasy to cross bridge over river?	24.0	23.0	1.044
Uneasy to cross wide street or square?	17.4		
Frightened by lightning?	46.0	22.0	2.091
Frightened in middle of night?	29.0	30.0	.967
Do you have great fear of fire?	55.0	39.0	1.410
<i>Abnormal anger</i>			
Speak on spur of moment, regret later?	84.1	13.0	6.470
<i>Morbid depression</i>			
Often feel just miserable?	54.0	12.0	4.500
Frequently burdened by sense of remorse?	64.0	23.0	2.785
Frequently in low spirits?	36.0	17.0	2.118
<i>Abnormal feeling</i>			
Are you a "crank" about food?	31.0	46.0	.673
Can you stand disgusting smells?	80.0	78.0	1.026
Are your feelings easily hurt?	75.3	65.0	1.158
Can you stand pain quietly?	31.0	33.0	.933
Can you stand sight of blood?	25.0	45.0	.556

TABLE 1 (*continued*)

Questions	Delinquent girls Percentage	Unselected girls Percentage	Ratio of neurotic responses of delin- quents to those of non-delinquents (unselected girls)
<i>Abnormal impulses</i>			
Ever been afraid might jump off when on high place?	35.0	21.0	1.667
Strong desire to commit suicide?	26.0	1.0	26.000
<i>Peculiar movements</i>			
Can sit still without fidgeting?	24.0	26.0	.923
Habit of twitching face, neck, or shoulders?	27.0	11.0	2.454
<i>Motor-coordination</i>			
Ever had habit of stuttering?	22.0	9.0	2.444
<i>Physical ailments</i>			
Things swim or get misty before eyes?	54.0	47.0	1.150
Often have pains in parts of body?	38.0	47.0	.808
Have feeling of suffocating?	36.0	23.0	1.565
Usually sleep well?	16.0	7.0	2.286
Feel well rested in morning?	32.6	18.0	1.811
Eyes often pain you?	35.0	23.0	1.522
Have great many bad headaches?	22.4	28.0	.800
<i>Fatigue and ennui</i>			
Feel tired most of time?	12.0	26.0	.461
Enjoy social gatherings to be with people?	39.0	19.0	2.053
<i>Phantasies</i>			
Ever bothered by feeling that things are not real?	55.0	27.0	2.036
Day-dreams improbable occur- rences?	62.0	30.0	2.067
<i>Ideas of persecution</i>			
At night troubled by idea that somebody is following you?	27.0	37.0	.730
Feel you are not satisfactorily adjusted?	47.0	16.0	2.937
Family treated you right?	48.0	1.0	48.000
<i>Indecision</i>			
Often find you cannot make mind up until time for action gone?	46.0	26.0	1.769

It is interesting to observe that the delinquent girl is less abnormal in feeling reactions than unselected girls, that is, she can stand pain more quietly, can stand unpleasant odors and the sight of blood more easily, and is less "cranky" about food.

"From the results of this table we would conclude that the delinquent girls in the penal institutions in North Carolina differ from Miss Mathew's group of unselected girls in their stronger desires (*a*) to be alone a great deal, and (*b*) to run away from home; (*c*) in their marked feelings of temper, (*d*) misery, and (*e*) persecution; and (*f*) their abnormal impulse to commit suicide" (2, p. 213).

Ten questions of the Neurotic Inventory relate rather specifically to the subject's sex life. These questions along with the percentages for the first and fourth quartiles are presented in Table 2. The data from these questions may be summarized as follows: The group as a whole is not ignorant of, and is not shocked by, sexual topics; also, there is little difference in

TABLE 2
PERCENTAGES OF NEUROTIC RESPONSES BY THE TOTAL GROUP, UPPER QUARTILE
AND LOWER QUARTILE, TO THE QUESTIONS RELATING TO THE
SUBJECTS' SEX LIFE

Questions	Percentage	Percentage	
		W.A.* Q_1	M.A.† Q_4
Shocked by sexual topics?	22	15	40.0
Fear of sexual inferiority?	25	18	81.0
Limit friendships to own sex?	44	20	71.0
Ignorant of sex?	20	2	17.5
Cynical about members of opposite sex?	23	11	40.0
Conflict in nature between sex and morality?	51	45	87.0
Are you shy with boys?	26	14	45.0
Are you shy with girls?	21	13	30.4
Generally indifferent to opposite sex?	25	14	25.0

*W.A.—Well adjusted, the upper 25% (Q_1).

†M.A.—Maladjusted, the lower 25% (Q_4).

these respects between the upper and lower quartiles.¹ Answers to the question: *Have you ever been afraid that you were sexually inferior to other women?* seems to vary more between Q_1 and Q_4 than those to any other question. The maladjusted in this case appear to have a keener sense of sexual inferiority. There does seem to exist a conflict in the

¹The upper quartile refers to the one-fourth scoring the most stable on the neurotic inventory, while the lower quartile refers to the one-fourth (group) scoring most unstable on the neurotic inventory.

nature of these girls between sex and their social concept of morality; that, in spite of their sexual experience (as a group), they are shy with boys generally, and limit their friendship mostly to their own sex; that the maladjusted are more cynical about members of the opposite sex, though the delinquent girls as a whole are not.

The findings of this study, although not wholly conclusive, appear to indicate:

1. A preponderance of emotional instability on the part of delinquent white women.
2. Some characteristic symptomatic responses of these prisoners which were also found in a recent study with men prisoners.
3. Characteristic abnormal fears and impulses of the delinquent women.

REFERENCES

1. BALL, R. J. The general emotionality of the prisoner. *J. Appl. Psychol.*, 1931, 15, 436-461.
2. LADU, L. B. & GARRISON, K. C. A study of emotional instability and intelligence of women in the penal institutions of North Carolina. *J. Soc. Forces*, 1931, 10, 209-216.
3. MATHEWS, E. A study of emotional stability in children by means of a questionnaire. *J. Delinq.*, 1923, 8, 1-13.
4. THURSTONE, L. L., & THURSTONE, T. G. A neurotic inventory. *J. Soc. Psychol.*, 1930, 1, 3-39.

North Carolina State College
Raleigh, North Carolina

DIFFERENTIAL FACTORS IN THE TESTING OF PERSONALITY: I. SEX DIFFERENCES

ROSS STAGNER

The wide-spread belief in sex differences in intelligence has recently been discredited by scientific investigations. The professional world is discovering that women make excellent lawyers, doctors, politicians. Even in motor skills a recent study (6) indicates that many tests show no difference, and some favor women rather than men. Despite these findings, there remains a strong conviction that characteristic differences exist between the sexes, and that these differences are a matter of psychological importance. Recently it has been suggested that they belong in that vague classification known as "personality."

Are there significant differences in the personality traits of men and women? By this we mean, is the score derived from a given objective "personality test" for a particular individual a function in part of the sex of that person? Is there ground for the belief that women are more "emotional" than men? These questions are indubitably of theoretical

importance; they are also of significance in the interpretation of test scores.

The author holds no brief for sex differences. His general observation has been that differences within one sex group are far more imposing than the difference between the averages of the groups. He does, however, feel that this hypothetical variation must be studied, and either shown to exist, when it can be taken into account in testing procedures, or laid by the heels as of no practical import.

That the difference, if it exists, has considerable significance, needs little demonstration. We must then develop separate percentile distributions for each group. We should be justified in inquiring into the source of the difference, whether inherited or acquired, and we might reasonably argue for a change in current educational practices, either in the direction of removing it (if acquired) or in adapting the curriculum to it. Methods of vocational prognosis would also be changed.

The present paper does not attempt to analyze all the data on sex differences in personality which are available. It is confined to certain tests with which the author has had personal experience: the Allport A-S Test; Colgate C 2; Neymann-Kohlstedt Test for Extroversion-Introversion; Social Intelligence; Thurstone Personality Schedule; Pressey X-O (Form A), and Hart's Social Attitudes Test. The chief attempt has been to combine data from several colleges, in an attempt to isolate sex differences from those due to the selective process which makes one college group differ from another. Since many studies have not presented the necessary statistical information, this attempt has met with somewhat inadequate success.

THE DATA

In Table 1 are summarized certain original data. The first six tests were taken by about 175 students at Gustavus Adolphus College, almost equally divided as to sex. The Pressey X-O Test was given to 154 students at the University of Wisconsin, who are also about equally divided.

From this table it appears that, while sex differences in personality may exist, they are in general not statistically reliable. The only critical ratio above the accepted minimum of three is that for the Thurstone Personality Schedule, which shows a reliable difference in favor of more neurotic scores for the girls. The critical ratio on the Hart Social Attitudes test [scored numerically as described in a previous publication (9)] is very nearly satisfactory. There are 99 chances in 100 that the men have more favorable attitudes toward social progress (as measured by this test).

These figures are open to certain criticisms. In the first place, the number of cases could well be increased. In the second, the group taking any given test was composed entirely of students from a single college. They were thus, in a sense, a selected group.

TABLE 1

Test		<i>N</i>	<i>M</i>	<i>S.D.</i>	<i>S.D.</i> <i>av.</i>	<i>D</i>	<i>S.D.</i> <i>D</i>	<i>C.R.</i>
Allport (ranks)	<i>M</i>	52	— .05	2.62	0.36			
	<i>F</i>	42	— 1.09	2.12	0.32	1.04	0.489	2.12
Laird C 2	<i>M</i>	84	15.35	4.27	0.46			
	<i>F</i>	91	16.47	5.13	0.53	1.11	0.711	1.56
Neymann-Kohlstedt	<i>M</i>	80	5.39	12.12	1.35			
	<i>F</i>	92	1.19	13.71	1.42	4.20	1.959	2.14
Social Intelligence	<i>M</i>	75	97.43	17.87	2.06			
	<i>F</i>	91	95.81	18.10	1.89	1.62	2.803	0.57
Thurstone	<i>M</i>	66	42.13	23.49	2.89			
	<i>F</i>	83	58.05	29.56	3.24	15.92	4.345	3.66
Hart Social Attitudes	<i>M</i>	75	31.71	6.42	0.74			
	<i>F</i>	83	28.69	6.98	0.76	3.02	1.066	2.83
Pressey X-O (Affectivity)	<i>M</i>	78	160.35	49.00	5.54			
	<i>F</i>	76	176.30	51.42	5.89	15.95	8.080	1.97

Garvey (4) has recently published a formula which enables us to combine the data of different investigators to give a single mean and standard deviation representing the group as a whole. We can thus combine published data on sex differences with our own to show what the results would have been if all these subjects had been in one group. Material that could be thus treated was available only for the Allport, Laird, and Thurstone tests. These are treated in Table 2.

A word is necessary about the statistical treatment of the Allport test. This is separately standardized for men and women, hence to speak of sex differences in the raw scores is an absurdity. We may, however, compare the two groups on the basis of the degrees of ascendance and submission as reported in the *Manual*. Giving these values from plus four (ascendance) to minus four (submission), we are enabled to treat them statistically. According to Table 2, then, both men and women are slightly submissive as judged by Allport's standards. We are inclined to doubt that the difference is anything but a matter of incomplete norms.

Table 2 includes the data of Table 1, and that of the following investigators: for the Allport test, Barrett, Jacobs, and White (2); C2, Broom (3) and Whitman (12); Thurstone, Allport (1), Barrett, Jacobs and White (2), Mears (7), and Thurstone and Thurstone (11).

DISCUSSION

Table 2 shows only one difference which meets the criterion of statistical reliability. This is for the Thurstone neurotic inventory. The Laird test does not give a sufficiently large critical ratio, although the total number of cases is somewhat above 1000. The Allport test likewise fails to meet the statistical requirements, and is, in addition, subject to the criticism voiced above, that this difference might be wiped out by enlarged norm values.

Comparing the results cited in Table 1, we find that on this smaller group the Thurstone test differentiated the sexes reliably; the Laird and Allport tests showed much smaller critical ratios than in Table 2. The increase in size of the ratio indicates that the direction of the difference is consistent in the combined groups.

Regarding the Social Intelligence Test, Hunt (5) has found women superior, but gives no statistical information. Our group of men is superior, but the critical ratio of .57 indicates that we may not attach great importance to the obtained difference.

On the other hand, the Hart Social Attitudes Test (a test of attitudes favoring reforms and social progress) seems to indicate a true difference between men and women, the men being somewhat more favorable to social improvement. This may perhaps be interpreted as supporting the traditional conservatism of the female. We should not attempt to defend this suggestion without further evidence.

The Neymann-Kohlstedt Test for Extroversion-Introversion is based on a theoretical attitude quite different from that of the Colgate Inventory, and the difference between men and women is larger for this than for the C2 Test, although in the same direction. Since definitions of introversion disagree among themselves so much, it is possible that men might be more extraverted by one criterion than by another. No statistical data for the Neymann-Kohlstedt Test have been furnished by its authors.

The results for the Pressey X-O Test are not easy to interpret. The

TABLE 2

Test		<i>N</i>	<i>M</i>	<i>S.D.</i>	<i>D</i>	<i>S.D. D</i>	<i>C.R.</i>
Allport (ranks)	<i>M</i>	108	— .002	2.34			
	<i>F</i>	102	— .860	2.16	0.85	0.304	2.79
Laird C 2	<i>M</i>	437	15.42	4.56			
	<i>F</i>	584	16.31	5.45	0.89	0.313	2.84
Thurstone	<i>M</i>	720	37.46	22.74			
	<i>F</i>	489	45.88	22.07	8.42	1.29	6.52

critical ratio here is less than two, yet it is large enough to strengthen belief that a true difference may exist in a larger group. The authors of the test believe that sex differences on the test may be important, since the differential units given in the *Manual* are different for the sexes.

We shall venture, in the light of these data, to sum up the question of sex differences for these tests as follows: a "true" sex difference exists in the case of the Thurstone neurotic inventory; for the Hart Social Attitudes and the Neymann-Kohlstedt Tests the probability of a true difference is rather high; for the Pressey and Laird Tests, doubtful; and for the Allport and Social Intelligence Tests, improbable.

Our suggestions for the use of these tests follow naturally from the above conclusions: for the Thurstone Test, separate frequency distributions should be developed [a distribution based on 720 men and 489 women is published elsewhere (10) for this purpose]. The Hart Test is now out of print, but users of similar tests should find sex differences a significant problem. Users of the Neymann-Kohlstedt Test could build up a frequency table [starting with one already published (10)]. For the other tests used, the indicated differences are too small to merit the labor of separate norms.

Attempts at explanation of sex differences are usually of little value. In the case of the neurotic tendency demonstrated by our figures for the Thurstone inventory, we believe that the explanation in terms of different training is adequate. In support of this we may cite several generally accepted facts: (1) the sex repression of women is still much more complete than that of men; (2) to an unfavorable environment more adaptive responses are available to the male than to the female child (as, e.g., running away, fighting back); (3) social disapproval of neurotic behavior is much less marked in the case of girls than of boys (this is especially true among the members of their own sex). A corollary of the third point bears on the test scores, though not on the neurotic temperament proper: Girls are more likely to confess neurotic traits than boys. The author's interviews with individual cases bear out this point noticeably.

Since the Thurstone Test correlates .53 with the C 2 inventory and —.34 with Neymann-Kohlstedt's Test (9), sex differences on the latter tests may in part be due to the overlapping of our current conceptions of introversion and neurosis. It is probable that further advances in testing must be made before these two aspects of personality can be separated for investigation.

SUMMARY

1. Sex differences are presented for seven well-known personality tests.
2. A test of neurotic temperament and one on attitudes toward social progress offer the only results approximating statistically significant sex differences.

3. It is believed that these differences are due to variance in the training of children.

REFERENCES

1. ALLPORT, G. W. Neurotic personality and traits of self-expression. *J. Soc. Psychol.*, 1930, **1**, 524-527.
2. BARRETT, D., JACOBS, L. B., & WHITE, E. M. Correlations between four personality tests and certain physical measurements. Unpublished thesis, Univ. Wis., 1931.
3. BROOM, M. E. Critical study of a test of introversion-extroversion traits. *J. Juven. Res.*, 1929, **13**, 104-123.
4. GARVEY, C. R. Sigmas of combined distributions calculated from sigmas, means and frequencies of component distributions. *J. Educ. Psychol.*, 1931, **22**, 307-310.
5. HUNT, T. Measurement of social intelligence. *J. Appl. Psychol.*, 1928, **12**, 317-334.
6. HUSBAND, R. W., & LUDDEN, M. J. Sex differences in motor skills. *J. Exper. Psychol.*, 1931, **14**, 414-422.
7. MEARS, C. E. Interrelationships between academic success, intelligence, and social adjustment. Unpublished thesis, Univ. Wis., 1931.
8. STAGNER, R. Emotional factors influencing retention. Unpublished thesis, Univ. Wis., 1920.
9. ———. Intercorrelations of some standardized personality tests. (In press.)
10. ———. Improved norms for four personality tests. (In press.)
11. THURSTONE, L. L., & THURSTONE, T. G. A neurotic inventory. *J. Soc. Psychol.*, 1930, **1**, 3-30.
12. WHITMAN, R. H. Sex and age differences in introversion-extroversion. *J. Abn. & Soc. Psychol.*, 1929, **24**, 207-211.

University of Wisconsin
Madison, Wisconsin

DIFFERENTIAL FACTORS IN THE TESTING OF PERSONALITY: II. DIFFERENCES IN MATURITY

ROSS STAGNER

In testing personality by the so-called objective methods which have recently been developed, we are likely to lose sight of the fact that human traits are plastic and dynamic. A test is essentially a cross-section of the individual at any given moment. In using this method we are therefore in danger of losing sight of that genetic aspect, which is one of the most important considerations in a study of personality.

Statistical methods are not entirely adequate to overcome this defect of the testing method. We may, however, determine the effects of maturity upon test scores and make allowances for these effects in the interpreta-

TABLE 1
AGE DIFFERENCES BY GROUPS
(Group I: ages 17-19; Group II: ages 21-30)

Test		<i>N</i>	<i>M</i>	<i>S.D.</i>	<i>D</i>	<i>S.D.</i> <i>D</i>	<i>C.R.</i>
Allport*	I	84	—,81	2.18			
	II	62	—,54	2.13	0.27	0.35	0.77
Laird	I	63	17.19	5.23			
	II	50	15.98	4.18	1.21	0.88	1.36
Neymann-Kohlstedt	I	66	—,24	11.35			
	II	49	+6.70	11.55	6.94	2.15	3.21
Social Intelligence	I	64	100.40	16.93			
	II	48	94.70	19.52	5.70	3.52	1.61
Thurstone*	I	92	54.08	25.25			
	II	85	49.00	23.41	5.08	3.64	1.39
Hart	I	66	29.23	7.66			
	II	50	31.56	6.18	2.33	1.28	1.82

*These items include the data of Barrett, Jacobs, and White (1).

tion of the score of a given individual. This procedure gives us an approximate correction for the maturity factor.

For the purposes of the present paper, maturity may be considered as being of two kinds: chronological maturity, as measured by the age of the individual; and educational maturity, determined in this case by the number of semesters of college attended. While these overlap to some extent, it is clear that in the average college group they will be almost independent variables. We have considered this fact sufficient justification for compilation of data on both points.

THE DATA

Six well-known tests of personality traits were used. These included: the Allport A-S Test; the Laird C 2 Inventory; Neymann-Kohlstedt Extroversion-Introversion Test; Social Intelligence; Thurstone Neurotic Inventory; and Hart's Social Attitudes Blank.

One hundred and seventy-five students in the sophomore, junior, and senior classes at Gustavus Adolphus College took some or all of the tests. The age range was from 17 to 30. For the age comparisons, scores were also available for over 100 students at the University of Wisconsin, on the Allport and Thurstone Tests, which had been administered by Barrett, Jacobs, and White (1).

Table 1 presents the group comparisons on age effects. For the Allport and Thurstone Tests, the two groups of subjects have been combined. For the others, only the Gustavus group is represented.

It is clear that in general the effect of age is not clear-cut. Only on one test (Neymann-Kohlstedt) does the critical ratio exceed the accepted minimum of three. The tendency here shown is for older students to be more extraverted, and this is confirmed by the results of the Laird Test, although in the latter case the difference is not reliable.

There is some evidence for the belief that the older students are better adjusted, as shown by the Thurstone Test. The Hart scores also seem to indicate that they are more favorable to social progress. On the Social Intelligence Test, they are surprisingly found to be below the younger group, a fact which no one would anticipate. The explanation is not, however, a matter of extreme difficulty. In this group, social intelligence correlated with general intelligence .45 (5), and it is a fairly universal finding that in a given educational level, the younger members are the brighter. We are not inclined to believe that social intelligence actually diminishes with age; the reverse is much more probably true.

On the Allport Test, the difference is too small to consider. This fact again goes counter to current opinion, and, while we might suggest explanations for the data, we prefer to wait for further findings on this question.

Table 2 presents product-moment correlations between test scores and chronological age, the trends shown confirming the data of Table 1.

Our results do not wholly conform to those of other investigators. On the Laird Test, for example, Whitman (6) found that a group of adults, both men and women, made higher scores than the corresponding college student groups. As these were alumni of the institution, they represent individuals probably of the same general intelligence level and social status. It may possibly be that the age curve for extraversion fluctuates considerably with time of life. Oliver (4) reports that cases from the tenth percentile (extraverts) are younger than those from the ninetieth

TABLE 2
CORRELATIONS OF TEST SCORES WITH CHRONOLOGICAL AGE

	<i>r</i>	<i>N</i>
Allport (ranks)	+.093	84
Laird C 2	— .081	156
Neymann-Kohlstedt	+.230	155
Social Intelligence	— .094	150
Thurstone	— .091	133
Hart Social Attitudes	+.172	158

(introverts), but the difference is not reliable. Flemming (2) agrees with the present study in finding increased extraversion in the older group.

On the Social Intelligence Test, the only data available are those of Hunt (3), which are presented in incomplete form and can be compared only impressionistically. An investigation of her table suggests that within a given educational level there is no correlation of scores with age. (An exception must be made of the high-school group, which shows a slight increase, probably due to the continuing growth of general intelligence.)

If we may sum up briefly the status of age differences, we shall state it as follows: first, extraversion probably increases with age up to a certain point, although later it may decrease; secondly, attitude toward social progress probably becomes more favorable with age; and, thirdly, that neurotic traits apparently diminish with age, although this result may have been due to greater frankness on the part of the younger subjects. Our data show quite conclusively that, if the age range is not considerable, differences due to this factor may reasonably be disregarded in the interpretation of scores. It would, however, be desirable to develop norms by age levels for tests of extraversion, it being necessary, first, to find a more valid test of extraversion.

Table 3 summarizes the results obtained when the data are grouped by college classes. Since the differences are mostly small and all unreliable, we have not attempted to present in tabular form the various possible comparisons.

In general, the same trends are shown in this table as in Table 1; that is, with increasing maturity, there occurs a rather pronounced rise in the degree of extraversion; there is a consistent diminution of scores on the neurotic inventory; the results from the other tests are inconsistent and difficult to interpret.

We may now inquire: if both age and education show certain consistent influences on personality test scores, is it possible to separate the two? A quantitative determination is impossible with the small group studied in this particular case. We may, however, compare the high and low values shown in Tables 1 and 3.

On the Laird test, Table 1 shows 17.19 and 15.98 as the two means; Table 3 shows 16.57 and 15.24 as the extreme values. We see that if only young people are counted, irrespective of their college class, a higher introvert score is obtained than when we take only sophomores, regardless of age. On the other hand, if we take only seniors, we get a higher extravert score than is found with the upper age group. We can infer from this that age and education are to some extent independent variables, and that in the present data the two are each influential in determination of scores. These speculations also hold for the Neymann-Kohlstedt Test.

The facts on the Thurstone Test seem somewhat different. Here the

TABLE 3
COMPARISONS OF SOPHOMORE, JUNIOR, AND SENIOR GROUPS

Test	Group	N	M	S.D.	S.D. av.
Allport (ranks)	Soph.	70	— .84	2.33	0.278
	Junior	16	+ .75	2.54	0.637
	Senior	6	— .50	2.32	0.949
Laird C 2	Soph.	78	16.57	4.67	0.529
	Junior	69	15.95	5.78	0.696
	Senior	24	15.24	4.01	0.818
Neymann-Kohlstedt	Soph.	81	+ 2.62	12.26	1.362
	Junior	67	+ 0.13	9.84	1.202
	Senior	22	+ 8.13	10.26	2.188
Thurstone	Soph.	63	57.94	27.45	3.459
	Junior	61	48.32	30.19	3.866
	Senior	22	45.86	22.45	4.787
Social Intelligence	Soph.	76	96.73	17.74	2.035
	Junior	70	97.35	16.66	1.991
	Senior	22	94.05	20.50	4.371
Hart	Soph.	69	29.13	7.13	0.858
	Junior	69	31.21	6.76	0.814
	Senior	20	30.30	5.60	1.253

age groups are much closer together (54.08 and 49.00) than are the class groups (sophomores, 57.94; seniors, 45.86). It does not, therefore, seem rash to assert that in the case of the Thurstone Test, education is an important factor in the determination of scores, and that in the present data the greater number of upperclassmen in the older group may have caused the age difference observed. If a large number of cases were available, we could compare age levels of the same class, but this is not possible at present.

Two possible explanations of this result suggest themselves: (1) the upperclassmen may be more sophisticated, hence cautious about giving information about themselves; (2) they may actually have acquired greater insight into their problems and solved some of them. Our observations of the students themselves does not bear out the latter suggestion, but favors the former only in a few cases.

SUMMARY

1. Maturity is a factor in the determination of scores on some personality tests. This is especially pronounced in the case of tests of extraversion and of neurotic temperament.

2. It is indicated that age is less important than education in the scores on a neurotic inventory.

3. Both factors are probably important in the determination of extraversion.

4. Other tests show differences which are not consistent and may depend on other factors.

REFERENCES

1. BARRETT, D., JACOBS, L. B., & WHITE, E. M. Correlations between four personality tests and certain physical measurements. Madison, Wis.: Univ. Wis. Library, 1931.
2. FLEMMING, E. G. Predictive value of certain tests of emotional stability as applied to college freshmen. *Arch. Psychol.*, 1928, No. 96. Pp. 61.
3. HUNT, T. Measurement of social intelligence. *J. Appl. Psychol.*, 1928, 12, 317-334.
4. OLIVER, R. A. C. Traits of introverts and extroverts. *J. Soc. Psychol.*, 1930, 1, 494-509.
5. STAGNER, R. Interrelation of some standardized personality tests. *J. Appl. Psychol.*, 1932 (in press).
6. WHITMAN, R. H. Sex and age differences in introversion-extroversion. *J. Abn. & Soc. Psychol.*, 1929, 24, 207-211.

*University of Wisconsin
Madison, Wisconsin*

THE MARRIAGE-AGE OF OLDEST SONS

W. D. COMMINS

It is part of the theory of analytical psychology that the emotional development of the child takes place by growing from within outward. In an early stage of his growth, the causes of adequate stimuli for his emotional reactions are states of his own organism to which he instinctively reacts in the way he does. He has not yet learned that there is anything in the universe besides his own precious self, so when he gives expression to anger, or fear, or love, he does not in this early stage, react to the mental impression of the person who is thwarting him, or of the object that is making the frightful noise or of the parent who is coddling him, but rather to the internal mental states that are aroused in a manner unknown to him. We might say, with some ambiguity, however, that he fears, and loves, and is angry at, himself and his own modifications, rather than that he fears, and loves, and is angry at, anything that is external and objective to him.

In the sphere of the love emotion, this stage is oftentimes referred to as the "auto-erotic" period. The child loves himself. He showers upon

himself the manifold attentions that he will later on come to direct toward other persons found about him. These other persons are at first his parents, then his playmates, and, finally, in an intensified and complicated way, a person of the opposite sex. The choice of a mate or lover is, as it were, the gathering together of a lot of scattered threads or tendencies and binding them together in a strong cord of attachment to the loved one. There are undoubtedly many influences emerging from the past life of a person to affect his choice of a mate.

In preparation for the passage to the "hetero-sexual" stage, the mother, so runs the theory, is a very important influence upon the boy (just as the father is supposed to be upon the girl), in the formation of his ideal of a love-mate. In a word, the mother is the boy's first sweetheart, and, when he comes in later life to make the choice of a second sweetheart, he will be influenced by the memory-image of the first, and he will choose one as near like his mother as possible. It would be a rather difficult thing to attempt to investigate this relation, statistically, from the viewpoint of character and temperament similarities, because we cannot very well reduce these things to numerical terms. It is an easier task when age is involved.

We might accumulate data upon ages with the following in mind: If the mother exerts an important influence upon the boy in his choice of a mate, by entering unconsciously into his ideal, then we might expect offhand that the oldest child in the family would marry a younger woman than the youngest child, because the oldest child would come to form his ideal of a mate when his mother would be in her early womanhood, and the youngest child would form his ideal when the mother would be more advanced in years. The two children would be influenced by the mother, if at all, at different ages in her career, the oldest child when she would be relatively young, and the youngest child when she would be relatively old. If this is true, then the oldest son should marry a younger woman, on the average, than the youngest son. To investigate this aspect of the problem, we should have to obtain the age of the women, when they married an oldest or a youngest son, and that, as we know, is not always an easy thing to do.

An alternative proposal would be to take the ages of the oldest and of the youngest sons of families at the time of their marriage to see if a difference would show up in this comparison. It might reasonably be supposed that the oldest sons would not only marry younger women, but would also marry at an earlier age in life than the younger sons, due to the same supposed influence of the mother-image. These data were at hand in the English *Who's Who*.

A tabulation was made of the ages at marriage of those who were mentioned as being the "eldest" sons in a family. A collection of 538 cases was made in about half the book. Instead of taking, for the other side of

the comparison, the ages at marriage of only the "youngest" sons of a family (because they were too few in number), the ages at marriage of all those who were recorded as being anything other than oldest sons were collected. Thus this second group included not only the youngest sons; but also the "second" or the "fourth" or a son of any relative position in the family with the exception of the first. Men whose relative position in their family was not mentioned were ignored. There were 537 of these "other-than-oldest" sons.

The following is the frequency distribution of ages at the time of marriage of the two groups:

	Age	Oldest	Other-than-oldest
Below	20	2	2
	21-25	79	62
	26-30	198	169
	31-35	122	142
	36-40	68	83
	41-45	38	53
	46-50	15	15
	51-55	6	9
Above	55	10	2
Total		538	537
Median		30.64	32.44
Difference	1.8	P.E. of difference	0.22

The median age of the oldest sons at the time of marriage is 30.64, while the median age of the other-than-oldest sons at the time of their marriage is 32.44. The difference, one of nearly two years, is reliable as a sample, if we may judge by its size in comparison with its probable error of 0.22. This would seem to indicate that the oldest sons, at least those who attain *Who's Who*, marry at an age that is almost two years earlier in life than the age at which their brothers marry.

It might be a rash judgment, and involve guilt of the logical fallacy of the "undistributed middle," to say that the younger age of the oldest sons at the time of marriage shows unmistakable evidence of the influence of the mother-image upon their choice of an ideal mate. The data are apparently in accord with that theory, but there may be other reasons as well to account for the facts. Some may be sociological, or economic, rather than psychological. It might be that in the case of most of those reported in the English *Who's Who* the right of primogeniture would prevail and this might account for some of the difference. If the oldest sons should inherit more than their brothers, then they might feel at ease economically rather

early in life, and so come to think of marriage, while their less-favored brothers might have to delay their marriage until they considered themselves competent to support themselves in an appropriate manner. It does not seem to be easy to rule out such a possibility. There may be others as well.

While there is nothing conclusive about the fact (and I believe we may say that it is a fact), that the oldest son marries at an earlier age on the average, than his brothers, it is at least suggestive of the possibility that the mother-image, according to the theory of analytical psychology, influences him in the choice of a mate, and accordingly the age at which he shall marry.

Catholic University of America
Washington, D. C.

BOOKS

ROBERT LATOU DICKINSON AND LURA BEAM. *A Thousand Marriages; a Medical Study of Sex Adjustment*. Baltimore: Williams & Wilkins, 1931. Pp. xxv+482. \$5.00.

A physician near the close of his working life summarizes his practice and seeks the help of a penetrating woman with a gift for saying things pregnantly in its interpretation. Further, he has been interested in the correlation of structure, pathology, function, and attitude, and his notes and the work based upon them amply reflect this interest.

What is the nature and significance of a monograph with such a background?

It is useful at this stage to omit from consideration the accident that the practice happened to be gynecological, and to confine oneself to the above generalized terms. So considered, the question really inquires about the place and value in human thinking of the entire medical methodology, of which the present work is an excellent example. Bull-session habits throw into gear the familiar Is-it-science language reflexes; but the more important problems seem to cluster about its accuracy and importance—that is, its probability, in the technical sense, and the richness of its cultural correlations. On the former point Miss Beam herself is enlightened and frank—the patient often doesn't know, she is often deliberately, and from her own viewpoint virtuously, untruthful; the doctor contributes his own biases in undetectable places; and the author paints in broad, bold sweeps on a canvas that must remain flat—and must not forget to blur the details here and there, using her own unverifiable judgment as to what will and what will not distort the picture, for the sake of preserving anonymity. What is the standard error of estimate of the final result—with what confidence may one draw conclusions from it regarding the next patient? If science be a matter of high probabilities, assuredly this is not science; it is more comparable with the extracted observations of a shrewd traveler making his careful way through a land closed to all but the most highly trained—which is not to diminish its tremendous importance, since the land is inaccessible and secrets of great value are hidden there.

Still keeping to the generalized topic, what are these valuable secrets which are open only to the gaze of the physician? They are, I think, of the nature of *Weltanschauungen*—the psychic results of the visitation of somatic pathology. Thus they have the same social and human value as does biography—when a new voyager must find a channel, a chart embodying the experiences of his predecessors, the methods they used in seeking a channel, and the kind and direction of channel found, is of the greatest importance.

The specific channel with which the present book is concerned was that of the satisfying conduct of the sexual life; the doctor was not the pilot; but the repairer—to be a bit Freudian—of stove-in bows and such; and his psychological interest was in discovering, in the course of the repair work, the individual principles of navigation in their functional interaction with the reefs encountered and the events of the log in general. We may very roughly indicate the nature of the results by stating that the Admiralty Office responsible for issuing bulletins to the public has been greatly misled—or misleading. Not that this is exactly news; but the reports from the wheel-house direct have a convincing definiteness that is both rare and significant as compared with even correct general advices on the general direction of the cruise. We had suspected, for example, that the sexual problem was of some consequence in widows, but had not been quite prepared for the stately dowager inducing five or six orgasms in the course of an evening at the concert. In general, perhaps, we may put this study beside those of Davis and Hamilton, arriving at the necessity of facing frankly (if we are to hold to any proposition concerning the undesirability of human misery) the fact that full and regular sexual satisfaction under conditions acceptable to the experiencer is as indispensable as an adequate supply of food—and probably especially for women. This is a large proposition; if accepted it would mean the death of a great many tabus which are especially dear to us; nor is the provision of appropriate engineering less problematic than the removal of the tabus; and yet—the proposition seems to be stubbornly there.

Of course there are a goodly number of specific findings of much value. Frigid and dyspareunic women, and women unhappy in marriage for any other reason, show a diminished fertility—which is perhaps not surprising, but clamors for consideration in connection with the troublesome problem of differential fertility; in which connection it was also found that women with long training of an educational or religious nature are subject to special difficulties in establishing marital functions—but may in some cases be helped by specific instructions given by a physician in whom they have full confidence. Also, passionate women are likely to be very catholic in their adventures—to the extent, for example, of 153 lovers distributed among seven women; possibly this also is not surprising—though one would like to discover, say by psychoanalytic methods, something of the psychic determinants of so ambitious a somatic hunger. The ghosts of our Puritan forefathers are still intruding unseen barriers between the bodies of young lovers—again not news, but poignant in the individual history. Autoerotism is so common as to be almost universal, homosexuality enters the records only slightly; the doctor presents conclusions without (as yet) evidence that the condition of the organs documents sexual practice somewhat. There is detailed material on fertility, contraception, sexual initiation of

brides (it is refreshing to learn that brides are still being sexually initiated), and limited material on the so-called perversions. The group as a whole is, of course, of distinctly selected status.

It is clear that the plucked fruit of the senior author's labors is a socially important document; it is only a little less clear that it is, from the larger point of view, a reconnaissance only.

RAYMOND R. WILLOUGHBY

Clark University
Worcester, Massachusetts

DOROTHY HAZELTINE YATES. *Psychological Racketeers*. Boston: Badger, 1932. Pp. 232. \$2.00.

In this book Dr. Yates has given the best account which has been made available to date regarding a species of pest which has been especially active during the last fifteen years—the psychological racketeer. Her book is based for the most part upon a first-hand study of the teachings, tricks, and advertising methods of a considerable number of lecturers on applied psychology—the sort who go from city to city holding short-stand runs for the purpose of developing personality, correcting mental attitudes, and giving insight and power to gullible business men and soulful, sentimental women. The book contains a large amount of interesting information, including excerpts from advertisements, the author's abstracts of lectures given, and life histories of some of the outstanding fakers in this field. That impostors of the type described in this book should be able to reap such a harvest in almost any city in the United States is not a little depressing to anyone who would like to have a modicum of faith in average human intelligence.

LEWIS M. TERMAN

Stanford University
California

BOOKS RECEIVED

BARZUN, J. The French race: theories of its origins and their social and political implications prior to the Revolution. (*Stud. Hist., Econ., & Pub. Law*, No. 375.) New York: Columbia Univ. Press, 1932. Pp. 275. \$4.25.

DICKINSON, R. L., & BRYANT, L. S. The control of conception: an illustrated medical manual. Baltimore, Md.: Williams & Wilkins, 1931. Pp. xii+290. \$4.50.

HUNTER, E. A sociological analysis of certain types of patriotism: a study of certain patriotic attitudes, particularly as these appear in peace-

time controversies in the United States. (Ph.D. thesis, Columbia Univ.)
New York: Author, 1932. Pp. 263. \$2.50.

PHELPS, H. A. Contemporary social problems. New York: Prentice-Hall,
1932. Pp. x+783.

REINHARDT, J. M., & DAVIES, G. R. Principles and methods of sociology.
New York: Prentice-Hall, 1932. Pp. xxvi+684. \$3.50.

SEARS, L. Responsibility: its development through punishment and reward.
New York: Columbia Univ. Press, 1932. Pp. xii+198. \$2.50.

VITELES, M. S. Industrial psychology. New York: Norton, 1932. Pp.
xviii+652. \$4.25.

SUBJECT INDEX

- Activity of young infants, relation of sex and race to, 118**
- Adolescence and developmental age, 469**
- Ambivalence, definition, 340**
- Association test and marriage selection, 311**
- Assortative mating, quantitative analysis, 297**
- Attitude scale construction and individual opinion, 283**
 - statements, suggested criteria for writing, 367
- Attitudes of college students, liberal vs. conservative, 320**
- Attractiveness, physical, racial judgments of, 463**
- Best friends, 385**
- Character measurement, problems of, 131**
- Children's questions concerning sex, 37**
- Church attendance and liberalism, 324**
- Class differences in ability to endure pain, 374**
- Cohesiveness of Thurstone Personality Schedule, 412**
- Conversation, functions of, 146**
- Crime and economic depression, 259**
- Criminality in children, 66**
 - test for, 189
- Cruelty-compassion, group test of, 189**
- Delinquency and economic depression, 259**
- Delinquents, social adjustments of, and inhibition of automatic responses, 91**
- Depression, economic, and crime, 259**
- Developmental age and adolescence, 469**
- Drunkennes and economic depression, 259**
- Economic depression and crime, 259**
 - Education and liberalism, 332
 - and marriage selection, 315
- Emotional characteristics of women prisoners, 472**
 - maturity, scale of, 3
 - responses and personality, 379
- Evolution, belief in, and liberalism, 325**
- Father's occupation and liberalism, 331**
- Friends, test ratings of, 385**
- Genius and nervous instability, 212**
- Homogamy, quantitative analysis, 297**
- Human sexual behavior, scientific study of, 161**
 - statistics on, 97
- Incentives, measurement of, 107**
- Inhibition of automatic responses and social adjustments of delinquents, 91**
- Intelligence and information of first-grade children, 362**
 - and language, 235
 - and liberalism, 323
 - and marriage selection, 309
 - and persistence, 83
 - sex differences in, 435
- Isolation of blocs in legislative body, 425**
- Juvenile delinquency and economic depression, 259**
- Language, influence on intelligence tests, 235**
- Liberalism in college students, 320**
- Marriage selection, quantitative analysis of, 297**
- Marriage-age of oldest sons, 487**
- Maturity differences and personality testing, 482**
 - emotional, scale of, 3
- Mental differences of the sexes, 436**
 - traits and marriage selection, 309
- Method of correction for guessing in true-false tests, 359**
- Michigan non-verbal series, 354**
- Natural tendencies and crime in children, 71**
- Nervous instability and genius, 212**
- Neurotic disposition and persistence, 83**
- Occupation and marriage selection, 315**

Opinion and contribution of attitude scale, 283

and marriage selection, 314

Persistence, a scale for measuring, 79

Personality measurement, problems of, 131

pleasing, 100

testing, differences in maturity, 482

sex differences in, 477

some aspects of, 376

Physical traits and marriage selection, 306

Pleasing personality, 100

Political affiliation and liberalism, 328

preference and marriage selection, 315

Psychoanalytic theory and marriage-age of oldest sons, 487

Psychophysiological analysis of sex differences, 434

Race and activity in young infants, 118

differences in ability to endure pain, 374

Racial judgments of physical attractiveness, 463

Reliability of emotional maturity scale, 20

of personality tests, problem of determining, 135

of scale for measuring persistence, 83

of test for liberalism, 323

Religious belief, change in, and liberalism, 327

denomination and liberalism, 324

Self-estimates of liberalism, 324

Sex and activity in young infants, 118

behavior, human, scientific study of, 161

some statistics on, 97

children's questions about, 37

differences in ability to endure pain, 374

in general intelligence, 435

in learning to walk a tight wire, 114

physiological explanation, 450

psychological or mental, 436

psychophysiological analysis of, 434

in testing personality, 477

Size of home community and liberalism, 331

Social adjustment and personality, 380

Sociological status and information of first-grade children, 362

Taboo, definition, 338

and totemism, aspects of, 337

Temperament and marriage selection, 313

Testing, attitude scale and individual opinion, 283

some aspects of personality, 376

Tests, analysis of sex differences in, 434

of emotional characteristics of women prisoners, 472

emotional maturity scale, 3

group measurement of cruelty-compassion, 189

of intelligence and language background, 235

measurements of incentive, 107

Michigan non-verbal series, 354

personality, sex differences, 477

maturity differences in, 482

problems of character and personality measuring, 131

a scale for measuring persistence, 79

suggested criteria for writing attitude statements, 367

true-false, method for correcting for guessing, 359

Thurstone personality schedule, suggested revision, 401

Thurstone neurotic inventory and women prisoners, 472

personality schedule, criterion of validity, 401

criticism, 240

properties of, 401

suggested revision, 401

Totemism, definition, 339

and taboo, aspects of, 337

Validity of attitude scale, 293

criterion of, for Thurstone personality schedule, 401

of cruelty-compassion test, 195

of emotional maturity scale, 20

of personality tests, problem of determining, 136

of scale for measuring persistence, 83

Women prisoners, emotional characteristics, 472